© 1965 by the American Psychological Association, Inc.

LANCASTER PRESS, INC., LANCASTER, PA.

CONTENTS OF VOLUME 70

Allison, J. Fixation Time as a Function of Stimulus Uncertainty 433
Anderson, N. H. Averaging versus Adding as a Stimulus-Combination Rule in Impression Formation 394
Arima, J. K. Human Probability Learning with Forced Training Trials and Certain and Uncertain Outcome Choice Trials 43
Banks, R. K. Effect of Pairing a Stimulus with Presentations of the UCS on the Extinction of an Avoidance Response in Humans 294
Banks, R. K., and Vogel-Sprott, M. Effect of Delayed Punishment on an Immediately Rewarded Response in Humans 357
Barclay, A. See Goulet, L. R.
Barzeele, J. See Bertelson, P.
Battig, W. F. See Brown, S. C.
Bell, C. R. Time Estimation and Increases in Body Temperature 232
Bertelson, P., and Barzeele, J. Interaction of Time-Uncertainty and Relative Signal Frequency in Determining Choice Reaction Time 448
Bevan, W., and Turner, E. D. Vigilance Performance with a Qualitative Shift in Reinforcers 83
Biederman, G. See D'Amato, M. R.
Bilodeau, I. McD. Transfer of Training across Target Sizes 135
Birch, D. Extended Training Extinction Effect under Massed and Spaced Extinction Trials 315
Black, R. W. Discrimination Learning as a Function of Varying Pairs of Sucrose Rewards 452
Black, R. W., and Spence, K. W. Effects of Intertrial Reinforcement on Resistance to Extinction Following Extended Training 559
Bogartz, R. S. Sequential Dependencies in Children's Probability Learning 365
Bourne, L. E., Jr. See Peterson, C. R.
Broen, W. E., Jr. See Nakamura, C. Y.
Brown, S. C., Battig, W. F., and Pearlstein, R. Effect of Successive Addition of Stimulus Elements on Paired-Associate Learning 87
Burstein, K. R. Effect of UCS Intensity upon the Acquisition of Conditioned Responses Acquired under a Lengthened Interstimulus Interval 147
Carlton, P. L. See Pavlik, W. B.
Casseday, J. H. See Di Lollo, V.
Ceraso, J., and Henderson, A. Unavailability and Associative Loss in RI and PI 300
Christ, R. E. See Seibel, R.
Cicala, G. A., and Corey, J. R. Running Speed in the Rat as a Function of Shock Level and Competing Responses 436
Coate, W. B., and Gardner, R. A. Sources of Transfer from Original Training to Discrimination Reversal 94
Collins, J. F. See Eriksen, C. W.
Corah, N. L. Effects of the Visual Field upon Perception of Change in Spatial Orientation 598
Corey, J. R. See Cicala, G. A.
Corso, J. F., and Levine, M. Sonic and Ultrasonic Equal-Loudness Contours 412
Craske, B. See Howard, I. P.
Creelman, C. D. Discriminability and Scaling of Linear Extent 192
Cross, D. V., Lane, H. L., and Sheppard, W. C. Identification and Discrimination Functions for a Visual Continuum and Their Relation to the Motor Theory of Speech Perception 63
Cross, K. See Trumbo, D.
Crovitz, H. F., and Schiffman, H. R. Visual Field and the Letter Span 218
D'Amato, M. R., Keller, D., and Biederman, G. Discriminated Avoidance Learning as a Function of Parameters of Discontinuous Shock 543
Davidson, J. R. See Martin, J. G.
Di Lollo, V., and Casseday, J. H. Graded Contrast Effects in the Judgment of Lifted Weights 234
Di Lollo, V., Ensminger, W. D., and Notterman, J. M. Response Force as a Function of Amount of Reinforcement 27
Doren, D. G. See Greenberg, M. G.
Dykman, R. A. See Galbrecht, C. R.
Ebenholtz, S. M. Positional Cues as Mediators in Discrimination Learning 176
Ekman, G., Hosman, J., and Lindström, B. Roughness, Smoothness, and Preference: A Study of Quantitative Relations in Individual Subjects 18
Elias, M. F. The Relation of Drive to Finger-Withdrawal Conditioning 109
Engen, T. Psychophysical Analysis of the Odor Intensity of Homologous Alcohols 611
Ensminger, W. D. See Di Lollo, V.
Epstein, S., and Fenz, W. D. Steepness of Approach and Avoidance Gradients in Humans as a Function of Experience: Theory and Experiment 1

Eriksen, C. W., and Collins, J. F. Reinterpretation of One Form of Backward and Forward Masking in Visual Perception 343
Eriksen, C. W. See Steffy, R. A.
Erlebacher, A. See Hill, W. F.
Feehrer, C. E. See Watkins, W. H.
Feinstein, S. H. See Rice, C. E.
Fenz, W. D. See Epstein, S.
Forsyth, G. A. See Newman, S. E.
Franchina, J. J. See McAllister, W. R.
Freeburne, C. M. See Weissman, S.
Fuhrmann, C. I. See Young, R. K.
Furth, H. G. See Youniss, J.
Galbrecht, C. R., Dykman, R. A., Reese, W. G., and Suzuki, T. Intrasession Adaptation and Intersession Extinction of the Components of the Orienting Response 585
Gardner, R. A. See Coate, W. B.
Gentile, A. See Peterson, L. R.
Gescheider, G. A. Cutaneous Sound Localization 617
Gogel, W. C. Size Cues and the Adjacency Principle 289
Goss, A. E. Paired-Associates Learning with Varying Relative Percentages of Occurrence of Alternative Response Members: Influence of Instructions 51
Gould, J. D., and Schaffer, A. Partial Visual Feedback of Component Motions as a Function of Difficulty of Motor Control 564
Goulet, L. R. Interlist Response Meaningfulness and Transfer Effects under the A-B, A-C Paradigm 264
Goulet, L. R., and Barclay, A. Comparison of Paired-Associate Transfer Effects between the A-B, C-A and A-B, B-C Paradigms 537
Greenberg, M. G., and Doren, D. G. Relationship between Latency and Remoteness in Preference Judgments 182
Hakes, D. T. See Young, R. K.
Hammond, K. R. See Peterson, C. R.
Harding, G. B. See Wickens, D. D.
Henderson, A. See Ceraso, J.
Hicks, R. Y. See Young, R. K.
Hill, W. F., Erlebacher, A., and Spear, N. E. Reminiscence and Forgetting in a Runway 201
Hill, W. F. See Spear, N. E.
Hohle, R. H. Detection of a Visual Signal with Low Background Noise: An Experimental Comparison of Two Theories 459
Holstein, S. B., and Premack, D. On the Different Effects of Random Reinforcement and Presolution Reversal on Human Concept Identification 335
Hosman, J. See Ekman, G.
Houston, J. P. Short-Term Retention of Verbal Units with Equated Degrees of Learning 75
Howard, I. P., Craske, B., and Templeton, W. B. Visuomotor Adaptation to Discordant Exafferent Stimulation 189
Hug, J. J. See Porter, J. J.
Jahnke, J. C. Primacy and Recency Effects in Serial-Position Curves of Immediate Recall 130
Jamison, S. M. See Peterson, M. J.
Johns, M. D. Transfer of a Pattern versus Component Discrimination Following Training in a Probabilistic Situation 506
Jones, R. L. See Martin, J. G.
Jung, J. Two Stages of Paired-Associate Learning as a Function of Intralist-Response Similarity (IRS) and Response Meaningfulness (M) 371
Karlin, L. Effects of Delay and Mode of Presentation of Extra Cues on Pursuit-Rotor Performance 438
Keller, D. See D'Amato, M. R.
Kimmel, E. See Kimmel, H. D.
Kimmel, H. D., and Kimmel, E. Sex Differences in Adaptation of the GSR under Repeated Applications of a Visual Stimulus 536
Kirwin, P. M. See Penney, R. K.
Koh, S. D. Scaling Musical Preferences 79
Konick, A. See Morin, R. E.
Koski, C. H., and Ross, L. E. Effects of Consummatory Response Punishment in Spatial-Discrimination Learning and Response Fixation 360
Kressel, K. See Winnick, W. A.
Lachman, R., and Tuttle, A. V. Approximations to English (AE) and Short-Term Memory: Construction or Storage? 386
Lane, H. L. See Cross, D. V.
Laughlin, P. R. Selection Strategies in Concept Attainment as a Function of Number of Persons and Stimulus Display 323
Levine, M. See Corso, J. F.
Lindström, B. See Ekman, G.
Lintz, L. M. See Little, K. B.

Little, K. B., and Lintz, L. M. Information and Certainty ... 428
Ludvigson, H. W. See McHose, J. H.
Martin, J. G., Davidson, J. R., and Williams, M. L. Grammatical Agreement and Set in Learning at Two Age Levels ... 570
Martin, J. G., and Jones, R. L. Size and Structure of Grammatical Units in Paired-Associate Learning at Two Age Levels ... 407
Marx, M. H. See McCoy, D. F.
Mayzner, M. S. See Tresselt, M. E.
McAllister, D. E. See McAllister, W. R.
McAllister, W. R., McAllister, D. E., and Franchina, J. J. Dependence of Equality Judgments upon the Temporal Interval between Stimulus Presentations ... 602
McCoy, D. F., and Marx, M. H. Competing Responses and the Partial-Reinforcement Effect ... 352
McHose, J. H., and Ludvigson, H. W. Role of Reward Magnitude and Incomplete Reduction of Reward Magnitude in the Frustration Effect ... 490
McPherson, S. See Morin, R. E.
Messick, D. M., and Rapoport, A. Expected Value and Response Uncertainty in Multiple-Choice Decision Behavior ... 224
Miller, A. J. See Peterson, C. R.
Morin, R. E., Konick, A., Troxell, N., and McPherson, S. Information and Reaction Time for "Naming" Responses ... 309
Muir, W. R. See Runquist, W. N.
Murdock, B. B., Jr. Signal-Detection Theory and Short-Term Memory ... 443
Myers, J. L. See Myers, N. A.
Myers, N. A., and Myers, J. L. A Test of a Discrimination Hypothesis of Secondary Reinforcement ... 98
Nakamura, C. Y., and Broen, W. E., Jr. Further Study of Effects of Low Drive States on Competing Responses ... 434
Newman, S. E., and Forsyth, G. A. Isolation Effects when Paired Associates are Presented Serially ... 334
Noble, M. See Trumbo, D.
Norman, D. A., and Wickelgren, W. A. Short-Term Recognition Memory for Single Digits and Pairs of Digits ... 479
Notterman, J. M. See Di Lollo, V.
Obrist, P. A., Wood, D. M., and Perez-Reyes, M. Heart Rate during Conditioning in Humans: Effects of UCS Intensity, Vagal Blockade, and Adrenergic Block of Vasomotor Activity ... 32
Olson, R. K. Generalization to Similar and Opposite Words ... 328
Pavlik, W. B., and Carlton, P. L. A Reversed Partial-Reinforcement Effect ... 417
Pearlstein, R. See Brown, S. C.
Penney, R. K., and Kirwin, P. M. Differential Adaptation of Anxious and Nonanxious Children in Instrumental Escape Conditioning ... 539
Perez-Reyes, M. See Obrist, P. A.
Peterson, C. R., Hammond, K. R., and Summers, D. A. Optimal Responding in Multiple-Cue Probability Learning ... 270
Peterson, C. R., and Miller, A. J. Sensitivity of Subjective Probability Revision ... 117
Peterson, C. R., Ulehla, Z. J., Miller, A. J., Bourne, L. E., Jr., and Stilson, D. W. Internal Consistency of Subjective Probabilities ... 526
Peterson, L. R., and Gentile, A. Proactive Interference as a Function of Time between Tests ... 473
Peterson, M. J., and Jamison, S. M. Effects of Distribution of Practice of the Acquisition Pairs upon Mediating Responses and Response Availability ... 549
Popp, R., and Voss, J. F. Acquisition-Transfer and Retention of S-R_1, and S-R_2 Associations as a Function of S-R_1, S-R_2 Pattern ... 304
Porter, J. J., and Hug, J. J. Effects of Number and Percentage of Rewarded Trials on the Acquisition and Extinction of Lever Pressing Using a Discrete-Trial Procedure 575
Posner, M. I., and Rossman, E. Effect of Size and Location of Informational Transforms upon Short-Term Retention ... 496
Postman, L. Unlearning under Conditions of Successive Interpolation ... 237
Premack, D. See Holstein, S. B.
Prokasy, W. F. Simulus Fluctuation, Reactive Inhibition, and Time between Trials in Classical Eyelid Conditioning ... 464
Rapoport, A. See Messick, D. M.
Reese, W. G. See Galbrecht, C. R.
Rhine, R. J. Preference for a Positive Evaluative Response in Concept Learning ... 632
Rice, C. E., Feinstein, S. H., and Schusterman, R. J. Echo-Detection Ability of the Blind: Size and Distance Factors ... 246
Rodwan, A. S. Coherence Detection: A Basic Mechanism ... 57
Rodwan, A. S. Primacy of a Form Criterion in Perceptual Judgments ... 231
Ross, L. E. See Koski, C. H.
Rossman, E. See Posner, M. I.
Runquist, W. N., and Muir, W. R. Intrasession Decrements in the Performance of the Classically Conditioned Eyelid Reflex ... 520

SCHAFFER, A. See GOULD, J. D.
SCHIFFMAN, H. R. See CROVITZ, H. F.
SCHUMSKY, D. A. Use of Negative Slope Transformations of Knowledge of Results on a Simple Motor Response 534
SCHUSTERMAN, R. J. See RICE, C. E.
SEIBEL, R., CHRIST, R. E., AND TEICHNER, W. H. Short-Term Memory under Work-Load Stress 154
SEKULER, R. W. Spatial and Temporal Determinants of Visual Backward Masking 401
SENGSTAKE, C. B. Perception of Deviations in Repetitive Patterns 210
SHAFFER, L. H. Choice Reaction with Variable S-R Mapping 284
SHEPPARD, W. C. See CROSS, D. V.
SPEAR, N. E., AND HILL, W. F. Adjustment to New Reward: Simultaneous- and Successive-Contrast Effects 510
SPEAR, N. E. See HILL, W. F.
SPENCE, K. W. See BLACK, R. W.
STEFFY, R. A., AND ERIKSEN, C. W. Short-Term, Perceptual-Recognition Memory for Tachistoscopically Presented Nonsense Forms 277
STILSON, D. W. See PETERSON, C. R.
SUMMERS, D. A. See PETERSON, C. R.
SUZUKI, T. See GALBRECHT, C. R.
TEICHNER, W. H. See SEIBEL, R.
TEMPLETON, W. B. See HOWARD, I. P.
THOMPSON, C. P. Effects of Interval between Successive Numbers and Patterns in Verbal Learning 626
TIGHE, L. S. Effect of Perceptual Pretraining on Reversal and Nonreversal Shifts 379
TIGHE, T. J. Effect of Overtraining on Reversal and Extradimensional Shifts 13
TRESSELT, M. E., AND MAYZNER, M. S. Anagram Solution Times: A Function of Individual Differences in Stored Digram Frequencies 606
TROXELL, N. See MORIN, R. E.
TRUMBO, D., NOBLE, M., CROSS, K., AND ULRICH, L. Task Predictability in the Organization, Acquisition, and Retention of Tracking Skill 252
TURNER, E. D. See BEVAN, W.
TUTTLE, A. V. See LACHMAN, R.
ULEHLA, Z. J. See PETERSON, C. R.
ULRICH, L. See TRUMBO, D.
UNDERWOOD, B. J. False Recognition Produced by Implicit Verbal Responses 122
VOGEL-SPROTT, M. See BANKS, R. K.
VOSS, J. F. See POPP, R.
WATKINS, W. H., AND FEEHRER, C. E. Acoustic Facilitation of Visual Detection 332
WEISSMAN, S., AND FREEBURNE, C. M. Relationship between Static and Dynamic Visual Acuity 141
WICKELGREN, W. A. Acoustic Similarity and Intrusion Errors in Short-Term Memory 102
WICKELGREN, W. A. See NORMAN, D. A.
WICKENS, D. D., AND HARDING, G. B. Effect of UCS Strength on GSR Conditioning: A Within-Subject Design 151
WILLIAMS, M. L. See MARTIN, J. G.
WINNICK, W. A., AND KRESSEL, K. Tachistoscopic Recognition Thresholds, Paired-Associate Learning, and Free Recall as a Function of Abstractness-Concreteness and Word Frequency 163
WOOD, D. M. See OBRIST, P. A.
WRIGHT, J. H. Test for a Learned Drive Based on the Hunger Drive 580
YOUNG, R. K., AND FUHRMANN, C. I. Function and Percentage of Occurrence of Response Members in Paired-Associate Learning 169
YOUNG, R. K., HAKES, D. T., AND HICKS, R. Y. Effects of List Length in the Ebbinghaus Derived-List Paradigm 338
YOUNISS, J., AND FURTH, H. G. Discrimination Shifts as a Function of Degree of Training in Children 424

Journal of
Experimental Psychology

VOL. 70, NO. 1 JULY 1965

STEEPNESS OF APPROACH AND AVOIDANCE GRADIENTS IN HUMANS AS A FUNCTION OF EXPERIENCE: THEORY AND EXPERIMENT[1]

SEYMOUR EPSTEIN AND WALTER D. FENZ[2]

University of Massachusetts

33 experienced and 33 novice parachutists rated their approach and avoidance feelings at different points in time preceding and following a parachute jump. For the novice parachutists, self-rated avoidance increased up to a point shortly before the jump, and then decreased. For the experienced parachutists, self-rated avoidance increased up to the morning of the jump, decreased to the jump, and increased after the jump. It was concluded that with experience the point of greatest anxiety is displaced backward in time. To account for this phenomenon, a miniature theory was presented which assumed that with continuous exposure to threat, 2 developments take place, a heightening of the gradient of anxiety and the development of a gradient of inhibition of anxiety, the latter having the steeper slope.

Sport parachuting provides an unusual opportunity for investigating approach-avoidance conflict in humans under intensely ego-involving circumstances. Observation of inexperienced parachutists suggests that for them the gradient of avoidance is steeper than the gradient of approach. The enthusiasm exhibited some time before the jump is seen to gradually wane and is replaced on the day of the jump with fear and uncertainty. Such is not the case with experienced jumpers, who appear to become increasingly eager to jump as the time of the anticipated jump approaches. This would suggest that for experienced jumpers the gradient of approach is steeper than that of avoidance. The present study attempts to verify the above observations by obtaining ratings of fear of, and desire to jump as a function of the sequence of events leading up to, and following, a jump.

METHOD

The *S*s consisted of 33 inexperienced and 33 experienced sport parachutists. The former had made between 1 and 5 jumps, and the latter over 100 jumps. Shortly after a jump, *S*s were asked to rate the strength of their approach and avoidance feelings at each of the following points in time: (1) last week,

---

[1] This paper was presented, in part, at the American Psychological Association Convention in 1963. The study was supported by Grant MH-01293 from the National Institute of Mental Health, United States Public Health Service.

[2] Now at the University of Waterloo.

(2) last night, (3) this morning, (4) upon reaching the airfield, (5) waiting to board the aircraft, (6) getting strapped, (7) boarding the aircraft, (8) during ascent in the plane, (9) at the "ready" signal, (10) stepping outside onto the step above the wheel of the plane, (11) waiting for the signal to jump, (12) freefalling, (13) after the chute has opened, (14) immediately after landing. Approach was defined as "looking forward to the jump, wanting to go ahead, being thrilled at the prospect of jumping"; avoidance as "wanting to turn back and call the jump off, questioning why you ever let yourself get into jumping, fear."

The *S*s were told to try to keep their ratings of approach and avoidance as independent from each other as possible. To this end it was pointed out that they could feel both frightened and intrigued with jumping at the same time, and that it was important in rating approach to consider only the attributes of approach, and to do the same for avoidance.

Following this, *S*s were instructed to select the point of strongest approach and assign it a value of 10, and then to select the point of weakest approach and assign it a value of 1. Having defined the range of their responses, they were told to rate each of the other points in a relative manner. They were instructed to use the same number more than once to designate points of equal intensity. Following ratings of approach, ratings of avoidance were made in a corresponding manner. It should be noted that the procedure for rating that was employed provides information only on the *relative* strength of approach or avoidance at the different points in time, and provides no basis for inferring absolute magnitude.

RESULTS

Novice Parachutists

In Fig. 1 it can be seen that self-ratings of approach and avoidance are, to a large extent, inverted replicas of each other. Apparently, it was not possible for *S*s to keep the two ratings independent. Self-ratings of approach are relatively high at Point 1, 1 wk.

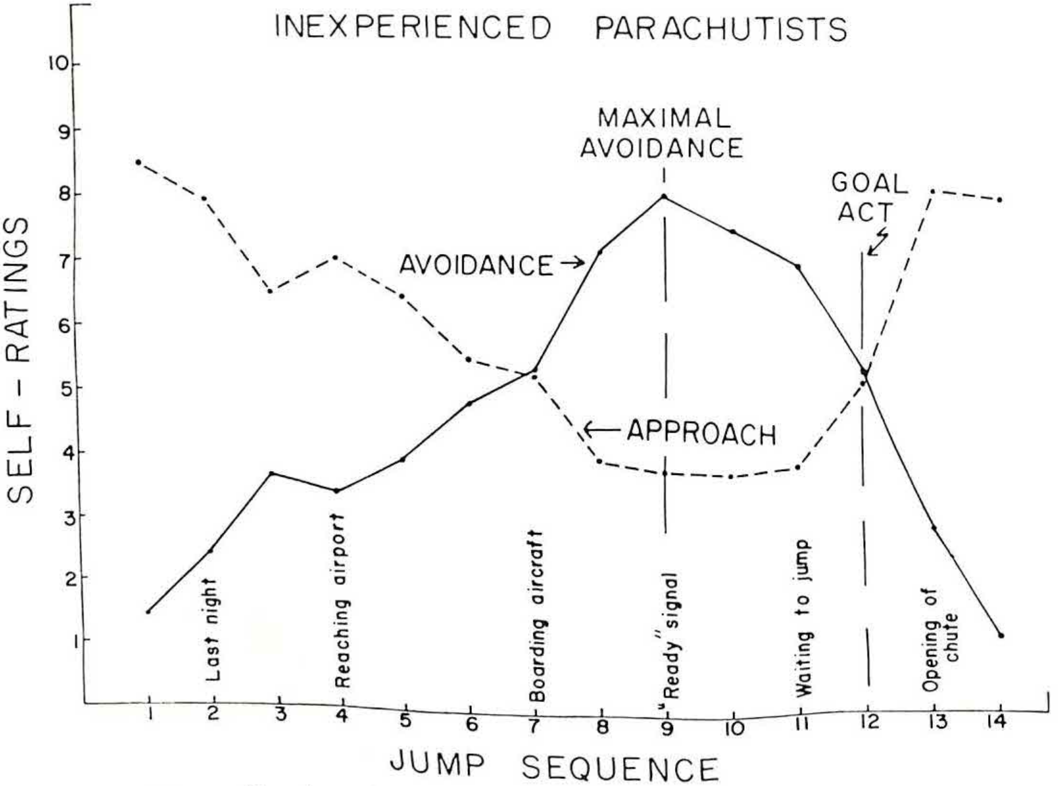


FIG. 1. Mean self-ratings of approach and avoidance for inexperienced parachutists as a function of a sequence of events leading to and following a jump.

before the jump, and fall off until shortly before the jump, after which they again rise. The converse holds for self-rated avoidance.[3] It is particularly noteworthy that the point of maximum avoidance corresponds to the ready signal at Point 9 and not to the goal act at Point 12, which is the moment of greatest danger, immediately preceding the opening of the chute.

An analysis of variance of the data represented in Fig. 1 would result in a spuriously high interaction between the time dimension and the approach and avoidance ratings, because the curves are largely equivalent but opposite forms of each other. An analysis of variance of each curve separately, on the other hand, would reduce reliability by eliminating half the data. A solution is provided by reversing one of the scales. Not only does this permit both sets of ratings to be pooled, but it allows for a determination of whether there is an interaction between the two scales and the time dimension. If such is the case, it establishes a degree of independence between the two kinds of ratings.[4]

Figure 2 presents curves for self-rated avoidance and for self-rated approach, the latter with the scale reversed, and referred to as "inverse approach." An *F* value of 41.39, significant at the .001 level, is obtained for the differences among the means of the pooled ratings of avoidance and inverse approach as a function of the time dimension (see Table 1). The failure of the *S*s × T interaction to reach significance indicates that the curve form is consistent among individuals except for deviations within chance expectancy. Of the 33 novice jumpers who took part in the experiment, 30 rate approach as greater a week before the jump than during the free fall, and corresponding results are obtained for self-ratings of avoidance.

Examination of the curves of avoidance and inverse approach in Fig. 2 reveals that they differ from each other in a systematic manner. From

---

[3] The observation that self-ratings of avoidance are greater than of approach at the time of the jump may appear to present a problem, as the parachutist, after all, does jump. The problem is more apparent than real. It will be recalled that the scales of approach and avoidance yielded relative ratings on a 10-point scale within each dimension, thereby requiring as many ratings of approach to be higher than avoidance as the reverse. Furthermore, as has been pointed out, the two kinds of ratings represent, for the most part, two ways of stating the same information. Accordingly, absolute differences between the two sets of ratings cannot be taken at face value.

[4] If *some* degree of independence is found, it does not, of course, invalidate the assumption that self-ratings of approach and avoidance represent, to a predominant degree, a single feeling state. Inspection of Fig. 1 and 2 makes it apparent that the two dimensions, pooled across *S*s, are almost perfectly inversely related.

It is important to keep in mind when evaluating the results of the analyses of variance that it is the *pooled* gradients of avoidance and inverse approach rather than the difference between them, i.e., the interaction, that is of primary interest. This is because self-ratings of approach and avoidance do not yield independent gradients of approach and avoidance, but are, by and large, two ways of reporting the same feeling, which, as will be seen later, can be viewed as a single net effect derived from independent hypothetical gradients of approach and avoidance. The major conclusions of the study would not be altered if only one of the sets of self-ratings were retained. This is not to say that differences between the curves of self-rated avoidance and inverse approach are of no importance, but to caution the reader against assuming that they are the major source of inference about the relative steepness of the hypothetical gradients. Such an error would be an easy one to make, as the forms of the curves of avoidance and *inverse approach* in Fig. 2 are similar in appearance to goal gradients of approach and avoidance as represented by Miller (1959).

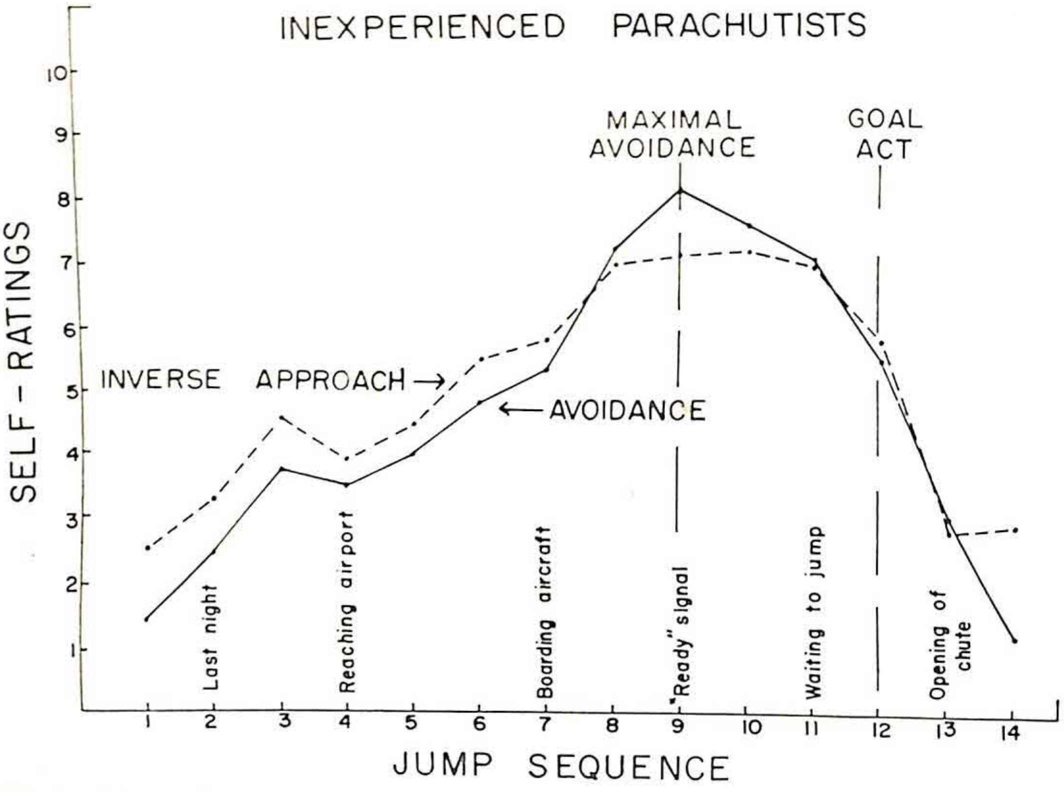


FIG. 2. Mean self-ratings of avoidance and inverse approach for inexperienced parachutists as a function of a sequence of events leading to and following a jump.

1 wk. before the jump (Point 1) to the ready signal (Point 9), the increase in self-rated avoidance is greater than the increase in inverse approach. From the ready signal on, the decrease in self-rated avoidance is greater than the decrease in inverse approach. The avoidance ratings, thus, tend to vary more sharply as a function of the time dimension than the approach ratings. This conclusion is supported by a significant A × T interaction in Table 1.

TABLE 1

ANALYSIS OF VARIANCE FOR INEXPERIENCED PARACHUTISTS ON SELF-RATINGS OF AVOIDANCE AND INVERSE APPROACH

Source	*df*	*MS*	*F*
Between *S*s	32	14.56	3.40***
Total within *S*s	891		
Avoidance vs. inverse approach (A)	1	23.71	1.85
*S*s × A	32	12.84[a]	3.13***
Time sequence (T)	13	252.04	41.39***
*S*s × T	416	6.09[a]	1.49
A × T	13	7.23	1.77*
*S*s × A × T	416	4.07[a]	

[a] Error term for *MS*s above it up to the next error term. The *S*s × A × T interaction was used to evaluate all other error terms.

* $p < .05$.

*** $p < .001$.

Separate analyses of variance for the ascending and descending portions of the curve in Fig. 2 indicate that in both cases the pooled gradients of avoidance and inverse approach are significantly related to the time dimension at the .01 level. While inspection reveals that avoidance tends to be steeper than approach for both portions of the curve, the interaction between kind of rating and time dimension is significant only for the ascending portion.

Experienced Parachutists

Figure 3, like Fig. 1, is divided at two points, one corresponding to the point of greatest avoidance, and the other to the goal act and point of greatest objective danger. There is an increase in self-rated avoidance (or conversely, a decrease in self-rated approach) up to the morning of the jump, followed by a more protracted decrease in self-rated avoidance (and a corresponding increase in self-rated approach) up to the goal act, and finally an increase in self-rated avoidance (or decrease in self-rated approach) after the goal act. Opposite to the findings for novice parachutists, self-rated approach is higher (and self-rated avoidance correspondingly lower) at the time of the jump than early in the sequence. This phenomenon is associated with a shift in the peak of self-rated avoidance for the experienced jumpers to an earlier point on the time dimension than for the inexperienced jumpers. It is noteworthy that after the goal act, the experienced jumpers report an increase in avoidance.

Figure 4 presents the curves of self-ratings of avoidance and inverse approach for experienced parachutists. An F value of 65.58, significant at the .0001 level (see Table 2), is obtained for the combined ratings as a function of the time dimension. Twenty-nine out of 33 experienced jumpers report that approach is greater during the free fall than on the morning of the jump, and corresponding results are

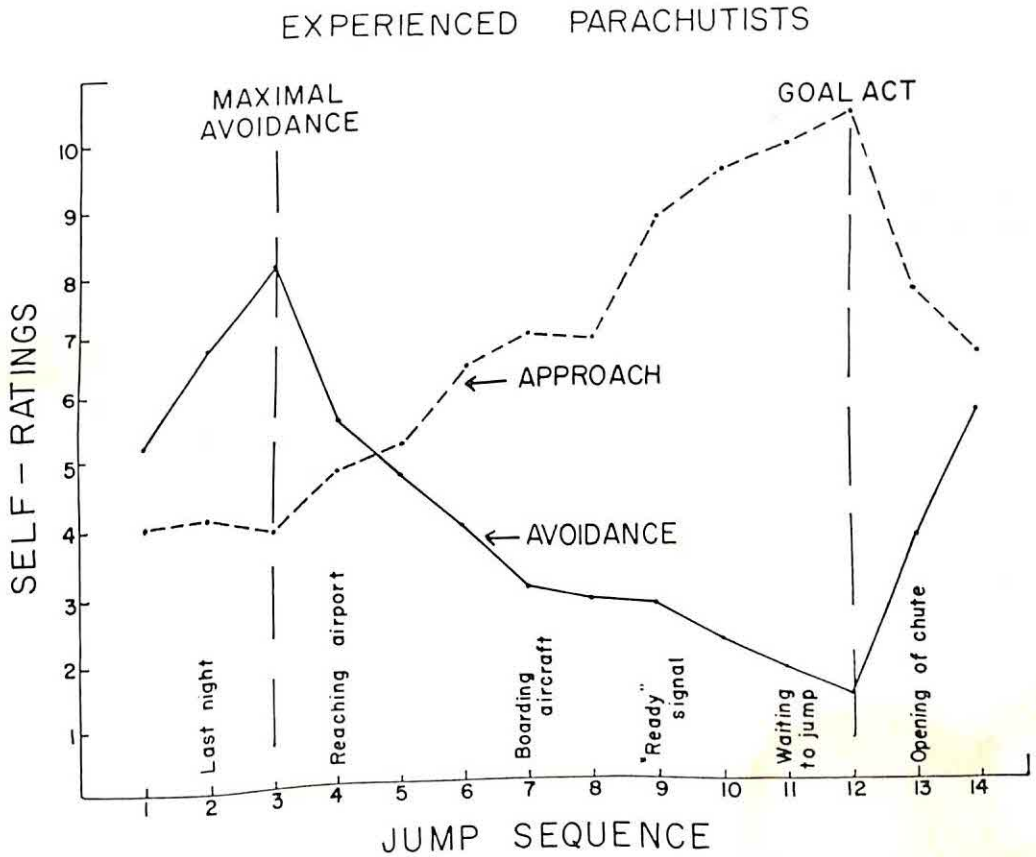


FIG. 3. Mean self-ratings of approach and avoidance for experienced parachutists as a function of a sequence of events leading to and following a jump.

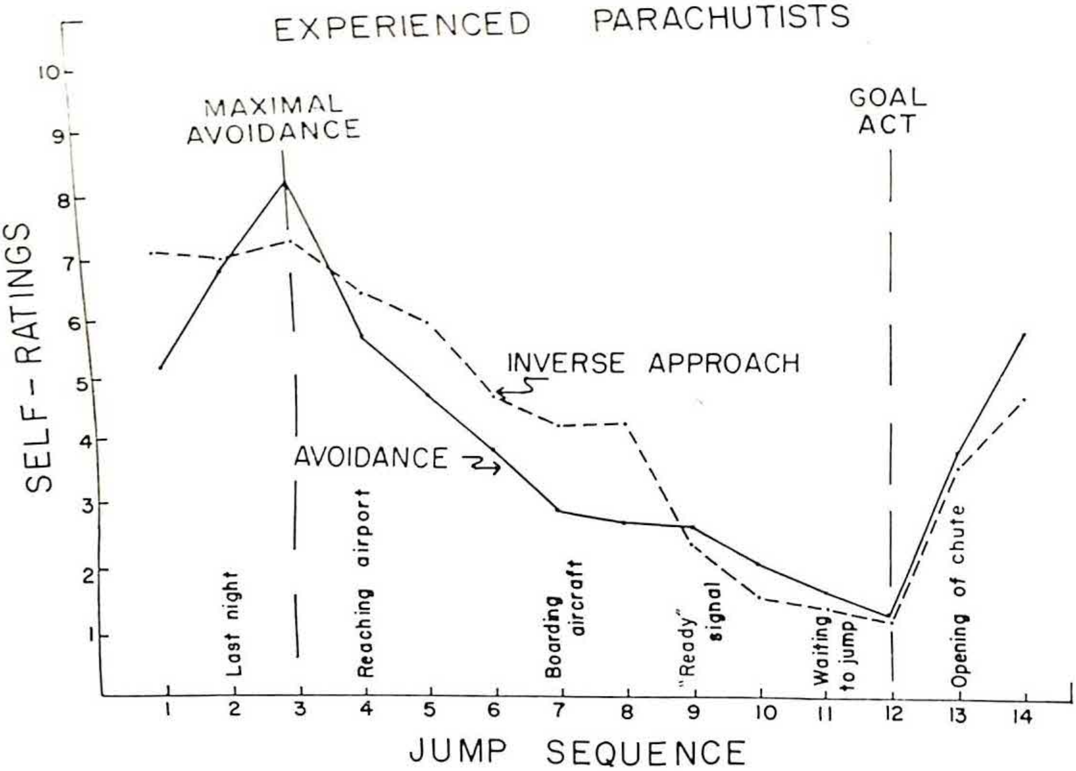


FIG. 4. Mean self-ratings of avoidance and inverse approach for experienced parachutists as a function of a sequence of events leading to and following a jump.

obtained for self-ratings of avoidance. Individual differences in curve form do not approach significance, as indicated by an *F* value of 0.93 for the *S*s × T interaction. The interaction between the time dimension and the two kinds of ratings is significant at the .01 level, indicating a degree of independence between approach and avoidance ratings.

TABLE 2

ANALYSIS OF VARIANCE FOR EXPERIENCED PARACHUTISTS ON SELF-RATINGS OF AVOIDANCE AND INVERSE APPROACH

Source	*df*	*MS*	*F*
Between *S*s	32	12.03	27.12***
Total within *S*s	891		
Avoidance vs. inverse approach (A)	1	24.03	1.65
*S*s × A	32	14.54[a]	3.29***
Time sequence (T)	13	265.14	63.58***
*S*s × T	416	4.17[a]	.93
A × T	13	13.92	3.14***
*S*s × A × T	416	4.43[a]	

[a] Error term for *MS*s above it up to the next error term. The *S*s × A × T interaction was used to evaluate all other error terms.

* $p < .05$.

*** $p < .001$.

Inspection of Fig. 4 indicates that in all three periods, the self-ratings of avoidance produce steeper gradients than inverse approach. An analysis of variance confined to the data preceding the peak of self-rated avoidance indicates that the pooled ratings of avoidance and inverse approach vary reliably over the time dimension, and that the slope of the gradient of avoidance is reliably steeper than the slope of the gradient of inverse approach. An analysis of the period from the peak of self-rated avoidance to the goal act produces corresponding results, although here both gradients are falling. Analysis of the postaction period indicates that while the rise in the pooled ratings of avoidance and inverse approach as a function of the time dimension is highly reliable, the two curves are not reliably different from each other. Nevertheless, as was the case for inexperienced jumpers, there is a tendency for avoidance ratings in all periods to vary more sharply as a function of time than inverse approach ratings, and, therefore, than approach ratings.

Table 3 summarizes the results of an analysis of variance in which inexperienced and experienced parachutists are both included. The main effect for groups is not significant, indicating that they do not differ in rating style. The interaction between groups and the time dimension yields an *F* value of 87.06, which indicates that the difference in the curve form for the two groups is highly reliable. At the same time, individual differences within groups in curve form (see source of variance *Ss* × T/G) do not approach significance, attesting to a high degree of uniformity in the form of the curve when experience is held constant.

TABLE 3

ANALYSIS OF VARIANCE FOR INEXPERIENCED AND EXPERIENCED PARACHUTISTS ON SELF-RATINGS OF AVOIDANCE AND INVERSE APPROACH

Source	*df*	*MS*	*F*
Total between *Ss*	65		
Groups (G)	1	161.91	12.17***
Ss/G	64	13.30[a]	3.13***
Total within *Ss*	1782		
Avoidance vs. inverse approach (A)	1	47.73	3.48
G × A	1	.001	
Ss × A/G	64	13.69[a]	3.22***
Time sequence (T)	13	70.58	13.76***
G × T	13	446.60	87.06***
A × T	13	10.87	2.56*
Ss × T/G	832	5.13[a]	1.21
G × A × T	13	10.28	2.42*
Ss × A × T/G	832	4.25[a]	

[a] Error term for mean squares above it up to the next error term. The *Ss* × A × T/G interaction was used to evaluate all other error terms.

* $p < .05$.

*** $p < .001$.

DISCUSSION

It was found that self-ratings of approach and avoidance, rather than representing independent processes, were largely inverted replicas of each other. As it is necessary to assume separate and opposed processes for conflict to exist, and as there is no question but that parachuting provides an acute conflict for the novice parachutist, it is necessary to infer hypothetical gradients of approach and avoidance of which the self-ratings are a net effect. According to Lewin's (1946) and Miller's (1959) models of conflict, independent gradients of approach and avoidance should both rise as a function of decreasing time to the goal, with the avoidance gradient rising more steeply than the approach gradient. The obtained results for the novice parachutists, with self-rated approach (or self-rated avoidance) treated as a net effect, are in accord with such an assumption, as in order for net approach to be high at a time remote from the jump and low at a time close to the jump, the hypothetical gradient of avoidance would have to be steeper than the hypothetical gradient of approach.[5] The hypothetical gradients as presented in Fig. 5 do not intersect, as the parachutist does jump, and it is necessary to represent approach as greater than avoidance at the goal. Figure 6 presents the net differences as derived to scale from Fig. 5. It is apparent that with

---

[5] The reasoning here is no different from that used to infer a steeper gradient of avoidance than approach when it is noted in a conflict situation that a child (Lewin, 1946) or rat (Miller, 1959) makes stronger approach responses at a distance than when close to a goal. If we may take the liberty of anthropomorphizing for a moment, Miller's rats, could they report their feelings in a manner consistent with their behavior, would reproduce the data of our parachutists.

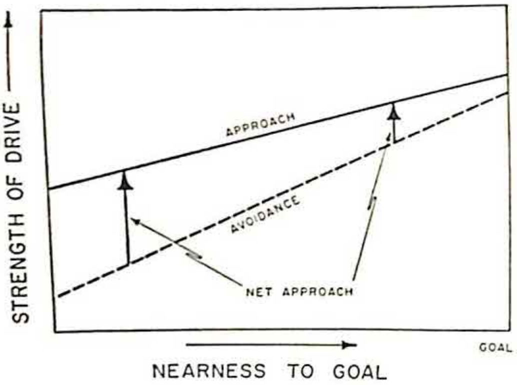


Fig. 5. Hypothetical gradients of approach and avoidance as a function of time to the goal in a conflict situation where the goal is reached in the absence of external restraint.[6]

increasing nearness to the goal, net strength of approach steadily diminishes. While the theoretical formulation represented by Fig. 5 and 6 adequately accounts for the general finding of a decrease in self-ratings of approach as the goal act is approached, it fails to account for the decrease in the avoidance ratings shortly before rather than immediately after the goal point. This phenomenon requires additional assumptions, which can best be presented in the discussion of the experienced parachutists.

For the experienced parachutists, self-ratings of approach are greater (or conversely, self-ratings of avoidance weaker) at the goal point than at a more remote time from the jump. Thus, for them, the data might appear to suggest that the hypothetical gradient of approach is steeper than the hypothetical gradient of avoidance. Such a state of affairs would have obvious adaptive advantages, as anxiety would automatically alert the parachutist well in advance of the potentially dangerous event, and be relatively low at the time of the jump. However, there are several reasons for favoring an alternate interpretation which assumes that the hypothetical gradient of avoidance is steeper than that of approach, and adds further assumptions. One difficulty with the first explanation is that it cannot account for the highly reliable early rise in self-rated avoidance (or corresponding drop in self-rated approach) for the experienced parachutists before the more protracted decline (see Fig. 3 and 4) to the goal point. Secondly, the assumption that the hypothetical gradient of approach is steeper than that of avoidance is inconsistent with the findings on self-rated avoidance and inverse approach. In three out of five comparisons self-ratings of avoidance were reliably steeper than of inverse approach, and in the other two cases the results were in the same direction. Thus, while it was concluded that the two ratings measure predominantly one feeling, to the small extent that an independent component is involved it supports the conclusion

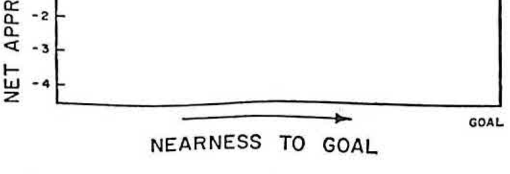


Fig. 6. Net approach strength as derived from hypothetical gradients of approach and avoidance in Fig. 5. (Units are arbitrary.)

---

[6] It will be noted that the ordinate is labeled as drive, rather than as strength of response or response disposition, as Miller labels it. Drive is defined as a directed force, with separate measurable components of direction and activation, the latter measurable by physiological indexes of arousal, such as the GSR. Drive, as so defined, is more closely related to Hull's (1943) excitatory potential than to his concept of drive, and is similar to Lewin's (1946) concept of vector. To derive total activation from hypothetical gradients of approach and avoidance, the two gradients are added disregarding algebraic sign. To derive net approach strength the gradients are added considering algebraic sign. (For a more detailed presentation of the theoretical approach, see Epstein, 1962.)

that the hypothetical gradient of avoidance is steeper than that of approach. Thirdly, the first explanation is of no use in accounting for the finding that the peak of self-ratings of avoidance for inexperienced jumpers precedes rather than coincides with the goal point, while the alternate theory is able to do so, using the same set of assumptions that account for the yet earlier peaking of the self-ratings of avoidance for experienced jumpers. Finally, the second theory has greater integrating power than the first in accounting for findings reported by others.

The considerably earlier peak of self-rated avoidance for the experienced than the inexperienced parachutists is consistent with findings in an earlier study of parachutists (Epstein, 1962; Epstein & Fenz, 1962) which used superficially different techniques, but of conceptual relevance to those in the present study. In the earlier work, GSR and reaction time were investigated as a function of a cue dimension of words related to parachuting embedded in a word-association test. On a control day, 2 wk. before a jump, experienced and novice parachutists could not be distinguished from each other, both producing monotonic gradients of similar slope on both dependent variables. On the day of a jump, however, experienced parachutists produced inverted V-shaped curves, while novices simply produced steeper monotonic gradients. With increasing experience, the peak of the inverted V was found to advance toward the remote end of the dimension. Tested repeatedly after different amounts of experience, all individuals demonstrated the same developmental pattern, although they differed in their rate of peaking back. As an example of an individual developmental pattern, Fig. 7 is presented. A theoretical explanation was offered which assumed that the down-turn on the inverted V is the consequence of an inhibitory process that serves to keep anxiety within nondisruptive limits. The progressive peaking back was accounted for by assuming that with repeated exposure to threat, two gradients develop, one based upon the generalization of conditioned anxiety, and the other upon learned inhibition of anxiety, the latter with steeper slope (see Epstein, 1962, for a more detailed presentation of the theory). In order to apply the theory to the present data, it is only necessary to substitute a time dimension for a cue dimension and ratings of fear and avoidance for a physiological measure. The finding that the peak of self-rated avoidance for novice parachutists precedes the goal act by a small amount can be explained by considering that the novices have had some experience, in training, in identifying with other jumpers, and in having made one to five jumps. Thus, the same process which produces a marked effect in the experienced jumpers, is presumed to produce a limited effect in the novices.

It should be noted that the theoretical formulation as presented above does not follow Hull's (1943) or Miller's (1959) views on conflict. One of the major differences is that we are postulating two levels of inhibition, one consisting of a hypothetical avoidance gradient which inhibits a hypothetical approach gradient, and the other of an inhibitory gradient which inhibits the hypothetical avoidance gradient. Thus, a concept of a hierarchy of inhibition is introduced in place of the single-factor inhibition theory of Hull and Miller. Such a conceptualization is necessary to account for the appearance of inverted V-shaped curves of avoidance in place of monotonic gradients. It is assumed that inhibitory gradients, as higher-order responses, are steeper than the gradients they inhibit, and that this relationship serves to free higher-order thought processes from primary stimulus generalization. This formulation differs from Miller's in that Miller assumes a dichotomy between higher- and lower-order processes, while the present position assumes a continuum based upon a chaining of inhibitory gradients of increasing slope.

It may be questioned as to whether it is indeed possible to inhibit anxiety. While the mechanism is not evident and deserves investigation in its own right,

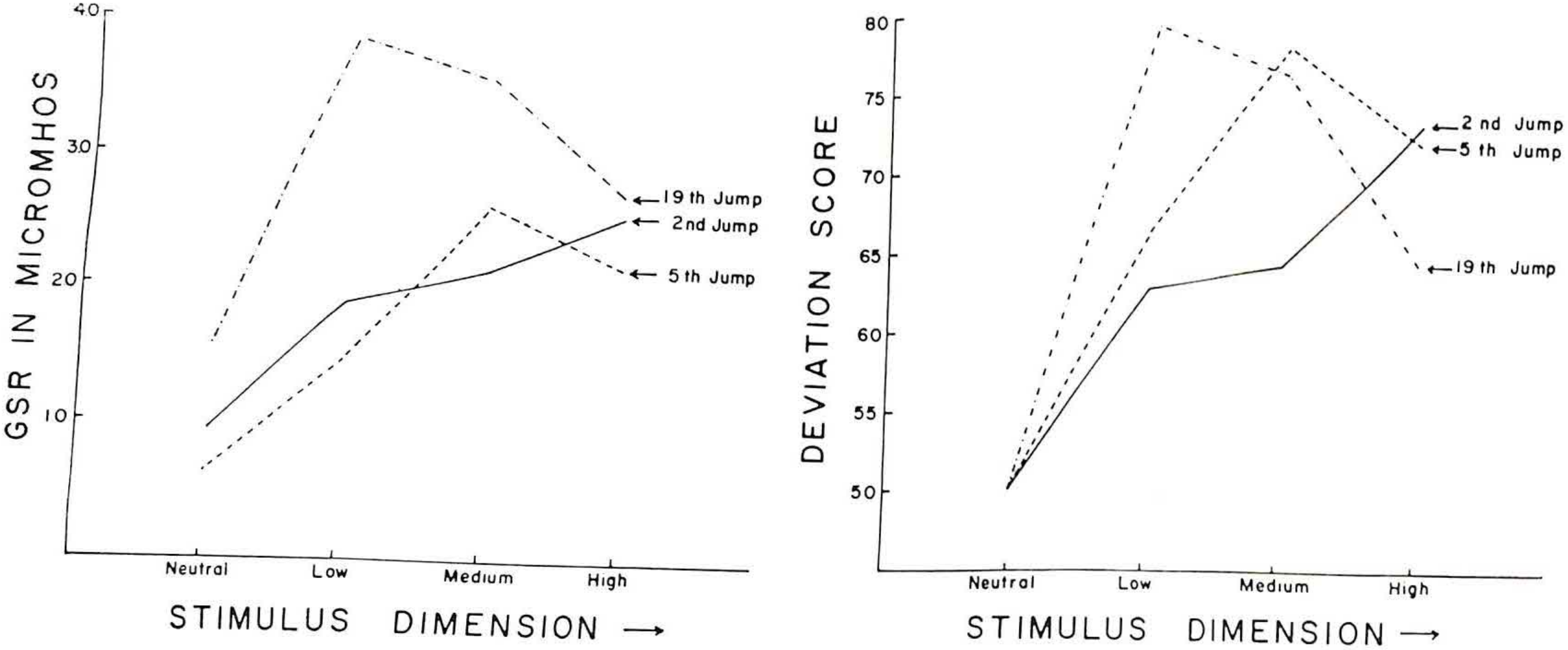


FIG. 7. GSR curves for a single *S* obtained at three levels of experience. ("Deviation score" refers to a transformation which controls for individual lability and mean reactivity as evaluated by responses to control stimuli [from Epstein, 1962].)

it is assumed that the inhibition of anxiety is similar to the inhibition of other autonomic responses which can often be delayed until an appropriate time and place for their expression are available. Such responses include vomiting, voiding, crying, and fainting from physical exhaustion. Very likely what is involved is the production of responses incompatible with the inhibited response, i.e., distracting responses, or the inhibition of goal-oriented responses, including those involved in perception, thought, and imagination. Whatever the mechanism, there is considerable evidence that anxiety can be inhibited. Thus, in the present study, the increase in self-rated avoidance following the goal act for the experienced, but not the novice, parachutists is consistent with the assumption that inhibited anxiety was released when inhibition no longer served a purpose. In this respect it will be recalled that the early peak of self-rated avoidance for the experienced jumpers was interpreted as resulting from a protracted period of inhibition before the goal act. A similar release of inhibited anxiety was noted in a study by Basowitz, Persky, Korchin, and Grinker (1955), where it was found that army paratroopers produced an unanticipated rise in anxiety some time after a stressful training period. In a study in progress on dreams we are finding that parachutists almost never report an anxiety dream immediately before a jump, but do so some time after the jump. In investigations of parachutists' reactions to word-association (Epstein & Fenz, 1962) and TAT-like tests (Fenz & Epstein, 1962) containing built-in stimulus dimensions, it was found that novice parachutists on the day of a jump failed to perceive anxiety-related words and explicitly denied fear of parachuting in stories to pictures of parachuting. At the same time, they showed an increase in anxiety responses to pictures unrelated to parachuting, suggesting displacement of inhibited anxiety.

There is reason to believe that the backward shift in the peak of avoidance with experience represents a fundamental adaptive mechanism in nonhuman as well as human animals. The phenomenon is remarkably similar to Pavlov's (Pavlov, 1927, 1928, 1941) "paradoxical phase." Pavlov noted that when his dogs were subjected to extreme states of stress over a prolonged period, a highly orderly sequence of changes took place in their conditioned response hierarchies. Following a period of total inhibition, a reverse gradient appeared, with the animal responding most strongly to stimuli that it had originally responded to least strongly, and least strongly to stimuli that it had originally responded to most strongly. Pavlov referred to this complete reversal as the "ultraparadoxical phase." Following this, there occurred a gradual shift along the stimulus dimension of the point of strongest response until the original gradient was restored. With repeated exposure to the same source of stress, the sequence unfolded itself more and more rapidly and finally disappeared. The Pavlovian phenomenon, as described above, is opposite in developmental direction to that observed in our parachutists. However, this is theoretically to be expected, as Pavlov attributed the above sequence of responses to the gradual dissipation of an inhibitory state, whereas in the theoretical formulation presented for the parachutists, the peaking back was assumed to be associated with increasing inhibition. Under conditions experimentally designed to increase inhibition, e.g., drowsiness and certain drug states, Pavlov was able to demonstrate a peaking back of the generalization gradient in the same direction as that found for our parachutists.

The shift in the peak of anxiety toward the remote end of the dimension, whether a cue dimension or a time dimension, has important implications for both adaptive and maladaptive behavior. On the adaptive side, it can serve to alert the person or animal in sufficient time for defensive reactions to be brought effectively into play. Secondly, the inhibitory reaction responsible for the inverted V serves to keep anxiety within relatively low and nondisruptive overall levels. It is presumably by this mechanism that the experienced parachutist is able to keep from experiencing the intense anxiety that the inexperienced parachutist must cope with. An interesting negative derivation related to this last advantage is that individuals who have successfully coped with stress over a prolonged period, if their controls ever were to break down, should be susceptible to greater panic and disorganization than inexperienced individuals. Such a derivation is actually supported by observations of wartime pilots (Bond, 1952). A further implication of the displacement of the peak of the inverted v is that while it should normally be adaptive in alerting the individual to an area of potential danger, should the conflict become unconscious, the displacement of the anxiety from its point of origin would result in what would appear to be meaningless anxiety, and the greater the time since the conflict, the more remote would the relationship become between the original conflict and the cues that elicit the anxiety. Space does not permit a full development of the implications of the inverted v-shaped curve, which will have to be reserved for a future theoretical paper.

REFERENCES

Basowitz, H., Persky, H., Korchin, S. J., & Grinker, R. R. *Anxiety and stress.* New York: McGraw-Hill, 1955.

Bond, D. D. *The love and fear of flying.* New York: International Univer. Press, 1952.

Epstein, S. The measurement of drive and conflict in humans: Theory and experiment. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1962.* Lincoln: Univer. Nebraska Press, 1962. Pp. 281–321.

Epstein, S., & Fenz, W. D. Theory and experiment on the measurement of approach-avoidance conflict. *J. abnorm. soc. Psychol.*, 1962, **64**, 97–112.

Fenz, W. D., & Epstein, S. Measurement of approach-avoidance conflict along a stimulus-dimension by a thematic apperception test. *J. Pers.*, 1962, **30**, 613–632.

Hull, C. L. *Principles of behavior.* New York: Appleton-Century-Crofts, 1943.

Lewin, K. Behavior and development as a function of the total situation. In L. Carmichael (Ed.), *Manual of child psychology.* New York: Wiley, 1946.

Miller, N. E. Liberalization of basic S-R concepts: Extensions to conflict behavior, motivation, and social learning. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. New York: McGraw-Hill, 1959.

Pavlov, I. P. *Conditioned reflexes.* (Trans. by G. V. Anrep) London: Oxford Univer. Press, 1927.

Pavlov, I. P. *Lectures on conditioned reflexes.* (Trans. by W. H. Gantt) New York: International, 1928.

Pavlov, I. P. *Conditioned reflexes and psychiatry.* (Trans. by W. H. Gantt) New York: International, 1941.

(Received April 17, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 1, 13–17

EFFECT OF OVERTRAINING ON REVERSAL AND EXTRADIMENSIONAL SHIFTS [1]

THOMAS J. TIGHE

Dartmouth College

32 rhesus monkeys were trained to criterion in a Wisconsin General Test Apparatus to respond to 1 dimension of a 2-dimensional discrimination task and then to respond either to the previously negative stimulus (reversal shift—RS) or to the previously irrelevant dimension (extradimensional shift—EDS). ½ of the Ss were given the discrimination shifts immediately upon reaching criterion in the original discrimination, while the other ½ received 200 training trials beyond criterion before undergoing the discrimination shifts. EDS was accomplished significantly faster than RS by both the criterion and the overtrained Ss. Overtraining had no effect.

Previous studies show that mature humans learn a reversal shift (RS), which involves a reversal of response to a previously relevant dimension, more rapidly than an extradimensional shift (EDS), which requires response to a previously irrelevant dimension (e.g., Buss, 1953; Harrow & Friedman, 1958; Kendler & D'Amato, 1955). But the contrary is true for young children (Kendler, Kendler, & Wells, 1960) and for the infrahuman organisms which have been tested (rats, Kelleher, 1956; chickens, Brookshire, Warren, & Ball, 1961; infant and mature rhesus monkeys, Tighe, 1964). However, the maturing child has been found to perform such tasks progressively more like the human adult (Kendler & Kendler, 1959; Kendler, Kendler, & Learnard, 1962).

Theoretical accounts of these age and species differences have stressed the role of mediating processes. Kendler and Kendler (1962) suggest that as children mature they become more likely to accomplish such problems on the basis of cues which are added to the stimuli of the task by mediating responses and which become the primary determiners of the discriminative response. The mediational sequence established in the initial discrimination remains appropriate in a subsequent RS whereas in an EDS it must be discarded and a new one formed. Consequently, an RS will be relatively easy for a mediator. The fact that young children and lower species find it relatively difficult to reverse is deduced from the assumption that these *S*s discriminate on the basis of single-stage S-R associations in the manner indicated by Spence (1936). For such *S*s the to-be-abandoned habit has to undergo more complete extinction in an RS than in an EDS. However, Kendler and Kendler have noted that appropriate response-produced stimulation may be developed in lower organisms by extended training (e.g., as presumably occurs in discrimination-reversal learning sets). Other forms of mediation theory which have been applied to discrimination-shift behavior assume that before *S*s can learn which stimulus to approach and which to avoid they must first acquire mediating responses which function to select out the relevant stimuli of the task. This mediating process has been conceptualized as the strengthening of observing responses (Zeaman & House, 1963) or the switching in of analyzers to the relevant dimension (Mackintosh, 1962). The more fully such processes are developed by training within a discrimination the more likely it is that *S*s will continue to respond to the relevant cues after an RS and consequently the faster will be the reversal learning. Concurrently, it becomes less likely that *S*s will respond in a subsequent EDS to the previously irrelevant but now relevant cues.

---

[1] This investigation was supported by Public Health Service Research Grant MH 07813-01, from the National Institute of Mental Health. The experiment was carried out at the Cornell University primate laboratory, which is supported in part by Grant M4516 from the National Institute of Mental Health. The author would like to thank Robert R. Zimmermann for the facilities and assistance afforded.

It follows from the latter formulations that infrahuman organisms will be likely to learn an RS more rapidly than an EDS following a high degree of overtraining (OT) in the original discrimination, or, at least, that OT will facilitate RS and retard EDS. According to single-stage S-R theory, however, OT should not facilitate RS. From the point of view of the Kendlers' formulation, then, the important consideration in predicting the effect of OT is whether or not this condition would result in the development of a cue-producing response which would enable rapid reversal. These issues are examined in the present experiment which studies the effect of extended training upon the shift performance of monkeys—a species which in view of its relatively advanced position on the evolutionary scale may be assumed to rank high among infrahuman species in the ability to develop and utilize mediating processes in problem-solving behavior.

METHOD

Subjects.—The *S*s were 32 rhesus monkeys ranging in age from 3 mo. to 16 mo. from the Cornell University Laboratory. The *S*s had prior discrimination training which was balanced over the conditions of this experiment. Each *S* received a daily ration of monkey chow and a vitamin sandwich.

Apparatus.—The animals were trained in two Wisconsin General Test Apparatus which were standard in their general structure. One, however, was of smaller scale to accommodate the younger *S*s. Mounted on the stimulus delivery trays were two sliding upright 3-in. square stimulus holders which covered food wells. The stimuli consisted of $\frac{3}{4}$-in. wide and $\frac{3}{8}$-in. thick plastic strips glued to a 3-in. square white plaque. The plastic strips varied in color, being red or green, and were arranged to form either a continuous $2\frac{1}{2}$-in. long vertical line or an interrupted vertical line of the same extent. Two such parallel lines separated by $\frac{1}{2}$ in. were centered on each plaque. Thus, on any one trial in original learning *S* had to choose between the stimuli in one of the following two pairs: (*a*) continuous red strips vs. interrupted green strips, or (*b*) interrupted red strips vs. continuous green strips.

Procedure.—All *S*s first learned a discrimination with one of the following as the positive stimulus: continuous strips (C), interrupted strips (I), red strips (R), or green strips (G). For example, if an *S* had C as the positive stimulus he was reinforced for choosing the continuous strips, regardless of whether they were red or green. Upon reaching criterion in the initial discrimination half of the *S*s (Criterion group) were immediately given the discrimination shifts. Half of these *S*s received a reversal shift (Group C-RS) and the other half received an extradimensional shift (Group C-EDS). For example, if an *S* had C as the positive stimulus in the initial discrimination he could then be given either an RS to I positive or an EDS to R or G positive. The remaining *S*s (Group OT) were similarly treated but only after having received 200 additional training trials on the initial discrimination during which criterion performance had to be maintained. This figure represents about 90% of the mean trials required to complete the original learning. An attempt was made to equate the major treatment groups in speed of learning the initial discrimination.

The general testing procedure followed the standard method used with the Wisconsin General Test Apparatus. The specific trial procedure consisted of *E* slowly moving the stimulus tray forward, pausing for 2 sec. just outside *S*'s reach, and then moving the stimuli closer to *S*. The *S* pushed back one of the two stimulus holders, revealing a raisin on correct choices. The *S*s received 25 non-

TABLE 1

MEDIAN NUMBER OF ERRORS AND TRIALS TO CRITERION DURING SHIFTS

Group	Relevant Dimension			
	Continuous-Interrupted		Red-Green	
	Errors	Trials	Errors	Trials
C-RS	233.5	550.5	225.5	237.5
OT-RS	345.0	650.0	98.5	216.5
C-EDS	118.5	312.5	35.0	112.5
OT-EDS	99.5	262.5	55.5	162.5

Note.—Each subgroup contains four *S*s.

TABLE 2

ANALYSIS OF VARIANCE OF TRANSFORMED TRIALS TO CRITERION IN DISCRIMINATION SHIFTS

Source	*df*	*MS*	*F*
Type of shift (S)	1	3.396	8.47**
Overtraining (OT)	1	.410	1.02
Dimension (D)	1	3.307	8.26**
S × OT	1	.114	
S × D	1	.001	
D × OT	1	.011	
S × OT × D	1	.146	
Within groups	23	.401	

** $p < .01$.

correction trials per day in each phase of the experiment, and were run to a criterion of 21 correct responses for each of 2 successive days both during original learning and for the postshift discrimination. Overtraining was given on consecutive days until 8 days of criterion performance had been achieved. Throughout training the order and position of stimuli followed an irregular sequence which provided that the same stimulus pair not appear more than twice in succession and that a correct stimulus not appear more than three times in succession on the same side. During the postshift discriminations the irrelevant dimension was held constant in order to minimize possible partial-reinforcement effects for *S*s undergoing the EDS. Each stimulus dimension was used equally often as the relevant dimension during original learning and the discrimination shifts. Each discrimination was begun on the day following completion of the preceding task.

RESULTS

The major treatment groups did not differ in speed of learning the initial discrimination ($F < 1$). All but three of the *S*s completed overtraining (OT) in 8 successive days. These three animals (two in the RS group and one in the EDS group) each required 1 additional day. During OT the mean number of errors per day was .82. The shift performance of the animals is summarized in Table 1. Trial and error scores for the discrimination shifts are presented in terms of the dimension involved since the I vs. C discrimination proved to be significantly more difficult than the R vs. G discrimination. Factorial analyses of variance were applied to trial and errors scores transformed to natural logarithms. Table 2 presents the analysis for the trial measure. Analysis of the error measure, which is not given here, provided conclusions identical to those obtained from the trial analysis, although the p values for type of shift and for dimension were $<.001$ and $<.05$,

TABLE 3

PERCENTAGE OF RESPONSES TO THE PREVIOUSLY POSITIVE STIMULUS IN RS FOR THE CRITERION AND OVERTRAINED GROUPS

Groups	Days in Shift							
	1	2	3	4	5	6	7	8
Criterion	83.0	61.5	58.0	52.5	48.6	46.3	45.2	43.6
Overtrained	92.0	79.4	69.7	56.6	54.8	48.0	50.7	43.2

Note.—The trends are the same for both the C-I and the R-G discriminations.

respectively, for the error measure. One animal in the OT-RS group had to be removed during the shift phase of the experiment due to illness, and the cell mean was substituted for his score. The mean number of errors before the first correct response following reversal shift was 25.4 for the overtrained group and 9.4 for the criterion group ($t = 1.4$, $p < .20$, two-tailed). Table 3 presents the percentage of responses to the previously positive stimulus on successive days in reversal shift for the criterion and overtrained groups.

DISCUSSION

The finding that monkeys learn an EDS more rapidly than an RS confirms a previous study with primates which employed different dimensions than those used here (Tighe, 1964). In that experiment, as in the present study, relative ease of shift was not affected by differences in the difficulty of the discrimination involved. The OT condition of the present experiment did not facilitate reversal behavior or retard extradimensional shift behavior regardless of which dimension was relevant. In fact, the measures of perseverative errors indicate that OT had a retarding effect on reversal performance. These results are in agreement with a single-stage S-R theory of discrimination learning. According to this view OT will have no effect on the relative ease of the two types of shift since additional training should merely result in proportionate increases in the habit strengths involved in the original discrimination. It also follows from this view that reversal behavior will be retarded by OT since the absolute strength of the original habit will be increased. However, the retarding effect on RS may be relatively slight since the strength of the original habit can be expected to be near asymptote with criterion learning. The present findings, then, are consistent with the assumption that *S*s discriminated on the basis of single-stage associations throughout the experiment.

These results do not agree with Mackintosh's (1962) study with rats which showed that OT in a brightness discrimination facilitated a subsequent reversal shift but retarded the learning of a horizontal-vertical discrimination. Although rats trained only to criterion learned the brightness reversal more slowly than the horizontal-vertical discrimination, these relations were reversed with OT. However, these data are not pertinent to the issue of whether OT will enable infrahuman organisms to learn an RS more rapidly than an EDS since the RS and EDS in this design involved different dimensions and thus preclude direct comparison of the relative ease of the two types of shift. But the question does arise as to why Mackintosh obtained an effect of OT while the present study did not. Consideration of procedural differences give rise to a hypothesis which might account for the discrepant results. First, in the reversal condition Mackintosh used a simultaneous brightness discrimination—the only situation in which OT has generally been found to facilitate reversal learning in animals. Secondly, a single dimension was present during OT in the Mackintosh study, while an additional varying irrelevant dimension was present during OT in the present experiment. It is possible, then, that these factors may have especially favored the development and attachment of mediating responses during OT. In other words, the perceptual properties of the task may play a critical role in determining the effectiveness of OT. This hypothesis receives some support from Tighe's (1963) experiment which employed shift procedures identical to those of the present study and found that 5- and 6-yr.-old children would accomplish an RS more rapidly than an EDS if they were first given nonreinforced pretraining designed to emphasize the independence and the dimensional nature of the properties of the task stimuli. Control *S*s in this experiment showed no difference in the relative ease of the two types of shift, a finding which confirms previous studies (Kendler & Kendler, 1959; Kendler et al, 1962).

The results of the present experiment, then, are consistent with the assumption that there is a fundamental difference in the manner in which human and infrahuman organisms accomplish discrimination shift problems, and indicate that this difference is not readily eliminated by extended training within the problem. But in order to resolve the question of whether or not these statements are generally applicable to the behavior of infrahuman species in such tasks, further evidence is required concerning the influence of variations in the testing procedures employed, particularly in regard to variations in the perceptual structuring of the task.

REFERENCES

Brookshire, K. H., Warren, J. M., & Ball, G. G. Reversal and transfer learning following overtraining in rat and chicken. *J. comp. physiol. Psychol.*, 1961, **54,** 98–102.

Buss, A. H. Rigidity as a function of reversal and nonreversal shifts in the learning of successive discriminations. *J. exp. Psychol.*, 1953, **45,** 75–81.

Harrow, M., & Friedman, G. B. Comparing reversal and nonreversal shifts in concept formation with partial reinforcement controlled. *J. exp. Psychol.*, 1958, **55,** 592–597.

Kelleher, R. I. Discrimination learning as a function of reversal and nonreversal shifts. *J. exp. Psychol.*, 1956, **51,** 379–384.

Kendler, H. H., & D'Amato, J. F. A comparison of reversal shifts and nonreversal shifts in human concept formation behavior. *J. exp. Psychol.*, 1955, **49,** 165–174.

Kendler, H. H., & Kendler, T. S. Vertical and horizontal processes in problem solving. *Psychol. Rev.*, 1962, **69,** 1–16.

Kendler, T. S., & Kendler, H. H. Reversal and nonreversal shifts in kindergarten children. *J. exp. Psychol.*, 1959, **58,** 56–60.

Kendler, T. S., Kendler, H. H., & Learnard, B. Mediated responses to size and brightness as a function of age. *Amer. J. Psychol.*, 1962, **75,** 571–586.

Kendler, T. S., Kendler, H. H., & Wells, D. Reversal and nonreversal shifts in nursery school children. *J. comp. physiol. Psychol.*, 1960, **53,** 83–88.

Mackintosh, N. J. The effects of overtraining on a reversal and a nonreversal shift. *J. comp. physiol. Psychol.*, 1962, **55,** 555–559.

Spence, K. W. The nature of discrimination learning in animals. *Psychol. Rev.*, 1936, **43,** 427–449.

Tighe, L. S. *Discrimination learning: Differentiation versus mediation.* (Doctoral dissertation, Cornell University) Ann Arbor, Mich.: University Microfilms, 1963, No. 64–3630.

Tighe, T. J. Reversal and nonreversal shifts in monkeys. *J. comp. physiol. Psychol.*, 1964, **58,** 324–326.

Zeaman, D., & House, B. J. The role of attention in retardate discrimination learning. In N. R. Ellis (Ed.), *Handbook of mental deficiency.* New York: McGraw-Hill, 1963. Pp. 159–223.

(Early publication received
September 21, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 1, 18–26

ROUGHNESS, SMOOTHNESS, AND PREFERENCE: A STUDY OF QUANTITATIVE RELATIONS IN INDIVIDUAL SUBJECTS[1]

GÖSTA EKMAN, JAN HOSMAN, AND BRITA LINDSTRÖM

University of Stockholm

A set of 7 surfaces were used as stimuli in a scaling experiment. For 10 Ss individual scales were obtained, by a method of ratio estimation, for (a) roughness, (b) smoothness, and (c) preference. The main results were: (a) For all Ss roughness was approximately a power function of the physical stimulus variable defined as the coefficient of friction. The exponent of the function varied greatly between Ss but exceeded 1 in almost all cases. (b) For nearly all Ss smoothness was approximately the inverse of roughness. (c) For nearly all Ss preference was directly proportional to smoothness.

It has been pointed out (e.g., Ekman, 1961b) that psychological mechanisms may sometimes turn out to be rather simple when studied at the proper level. The psychological variable of similarity has, thus, been demonstrated to be a simple function of measurable psychological aspects of the percepts entering into the comparison (Eisler, 1960; Eisler & Ekman, 1959; Ekman, Engen, Künnapas, & Lindman, 1964; Ekman, Goude, & Waern, 1961).

The present paper reports a study in a new series of investigations of psychological mechanisms, in which a preference variable is studied as a function of other psychological variables. This type of problem has been rather neglected; one of the few instances of a similar approach was a color preference study by Guilford (1939). The particular purposes of the present investigation were: (*a*) to investigate *preference* as a function of the subjective variables *roughness* and *smoothness*, (*b*) to investigate the subjective variables roughness and smoothness in relation to each other and to the corresponding stimulus variable, and (*c*) to study all these relations on the level of the individual *S*.

METHOD

Subjects.—Ten *S*s took part in the experiment, eight females and two males. All *S*s were students of psychology with some previous experience of ratio estimation.

Stimulus materials.—Seven stimulus surfaces were used. Five stimuli (No. 1–5) were pieces of sandpaper, one (No. 6) was a piece of cardboard, and one (No. 7) a piece of ordinary writing paper. Several identical stimulus sets were prepared and used successively in order to avoid any noticeable change in surface due to use.

Static coefficients of friction were experimentally determined in relation to a sample of finger tips under conditions similar to those of the main experiment. From the number of grits used in the production of the sandpapers an alternative stimulus measure, 1/(No. of grits), was obtained which can be regarded as proportional to the diameter of the grains. The latter type of stimulus characteristic was used by Stevens and Harris (1962) in their scaling of roughness and smoothness. The data of the present experiments (as well as additional experiments covering a wider stimulus range) indicate, however, that the two sets of stimulus measures are not proportional to each other. Under these circumstances the coefficient of friction was chosen as stimulus measure.

However, the physical attributes determining the degree of perceived roughness are probably not described exhaustively by the coefficient of friction. Since the sandpapers are most similar to each other in other physical respects, the relations between the stimulus variable and the subjective variables may be most adequately studied for Stimuli No. 1–5.

---

[1] This investigation was supported by the Swedish Council for Social Science Research. The work was performed while Jan Hosman, from the University of Leyden, held a scholarship at the University of Stockholm.

Experimental procedure.—The stimuli were presented in pairs of surfaces, each about 3 × 10 cm., covered by a screen in such a way that *S*, without seeing the surfaces, could comfortably stroke them with his two index fingers.

There were three experimental conditions: (*a*) roughness estimation, (*b*) smoothness estimation, and (*c*) preference estimation. Each *S* participated in all conditions, one condition in each of three sessions on different days. Right-left position was rotated for each *S*, and the order of conditions was rotated between *S*s. In each condition *S* judged two random orders of 42 pairs, i.e., each of the 21 different stimulus combinations was judged four times by each *S*.

In the *roughness*-estimation sessions *S*s were instructed as follows:

> You will be presented with pairs of sandpapers. Your task is to give a judgment about the roughness of the sandpapers. Take your time just to get a good idea how rough one paper is compared with the other. Give the judgment in free numbers. For instance, if the paper on the right seems to be three times as rough as the paper on the left, say: the ratio is three to one. You may also use numbers like 300 to 100, or 45, to 15, or 1 to 0.33.

In the *smoothness*-estimation sessions the same instruction was used with the words "smoothness" and "smooth" substituted for "roughness" and "rough."

In the *preference*-estimation sessions *S*s were instructed in the same way to give "a judgment about the character of the surface in terms of pleasantness or preference."

The "free" form of the ratio-estimation method used in these experiments is similar to the form suggested by Mashhour (1964).

RESULTS

Experimental data.—The pairs of numbers reported by *S* were transformed to ratios. Arithmetic means were computed for the four ratios obtained from each *S* for each stimulus combination in each condition. Individual scales were constructed from the ratio matrices according to a procedure described by one of the present authors (Ekman, 1958, Equation 14).

The psychophysical function.—The subjective scales of roughness and smoothness are plotted against the stimulus variable, in log-log coordinates, in Fig. 1, in which there is one diagram for each *S*. Straight lines represent the general trend of all data in first approximation. For most *S*s the fit would be considerably better (and the slope of the lines a little steeper) if only the sandpaper stimuli were included. (A slightly better fit may also be achieved by introducing an additive constant [Ekman, 1958, 1961a].) It can be concluded that power functions describe the psychophysical relation in reasonably good approximation, at least for the sandpaper stimuli, for all *S*s. The actual fit is most adequately judged in Fig. 2, where the roughness functions are shown, for all seven stimuli, in linear coordinates; the goodness of fit of the smoothness functions is much the same.

This result is of considerable interest, for very little is known so far about the psychophysical function for individual *S*s. Almost all previous conclusions have been based on group data. One investigation by Pradhan and Hoffman (1963) was in part concerned with this problem. According to these authors, the power function did not describe the majority of individual data in experiments with weight lifting, although it did represent the group data fairly well; hence, it was concluded that the power function might be an artifact of grouping. On the other hand, recent investigations have shown a very satisfactory fit of power functions to individual loudness data (Stevens & Guirao, 1964) as well as to data

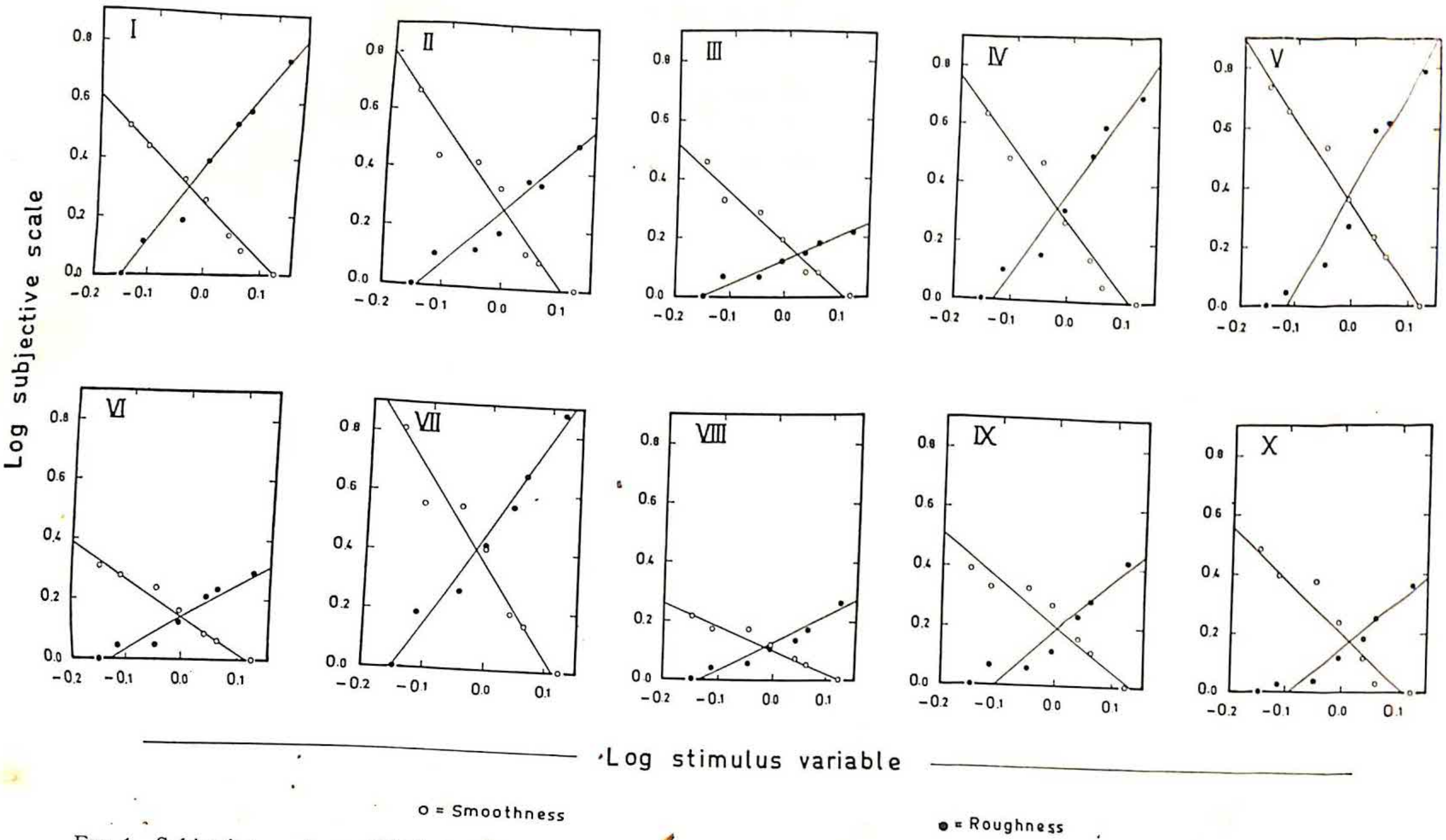


FIG. 1. Subjective roughness (filled circles) and subjective smoothness (empty circles) plotted against the coefficient of friction, in log-log coordinates. (Each graph represents an individual *S*.)

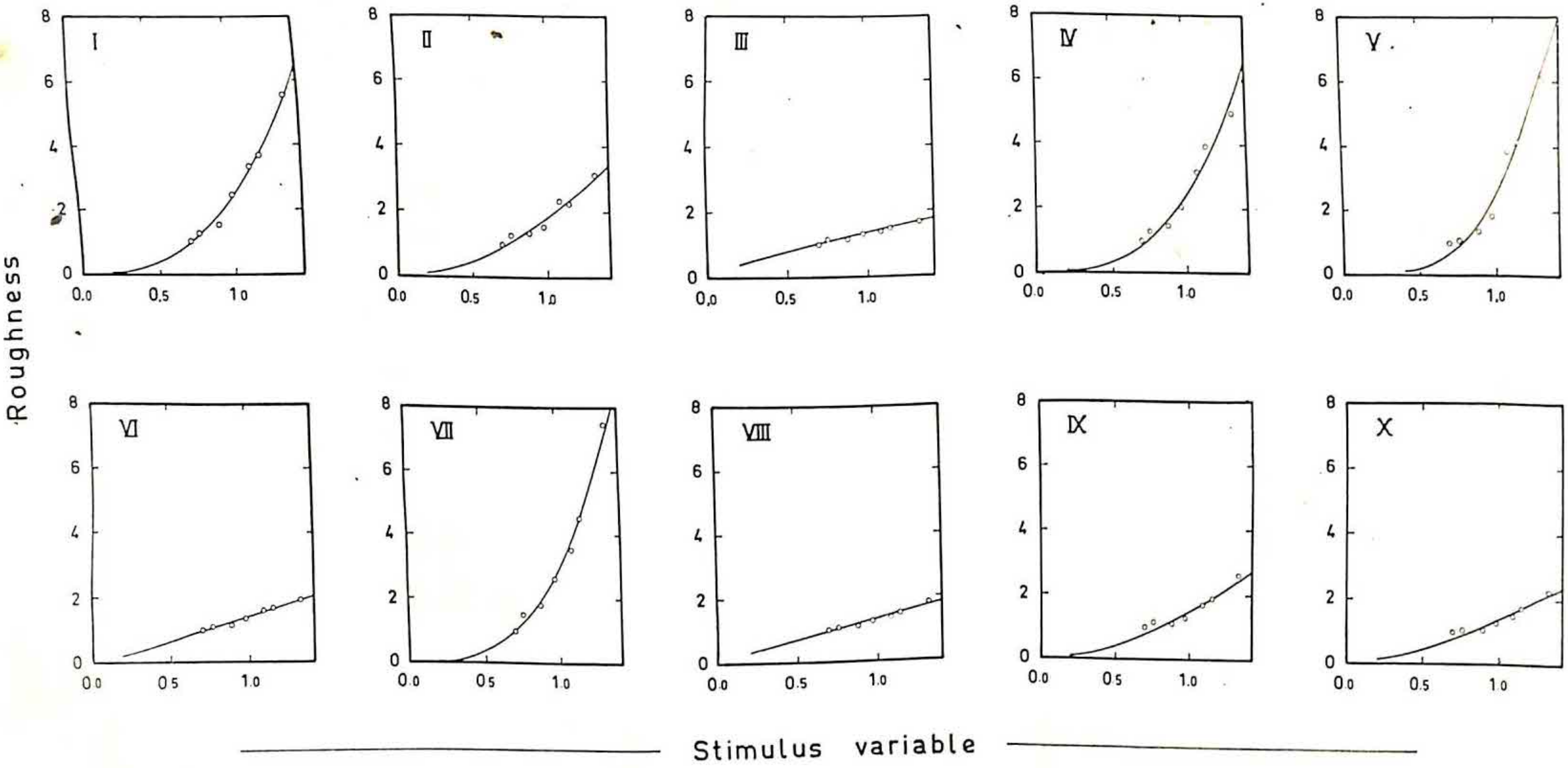


FIG. 2. Subjective roughness plotted against the coefficient of friction, in linear coordinates. (Each graph represents an individual *S*.)

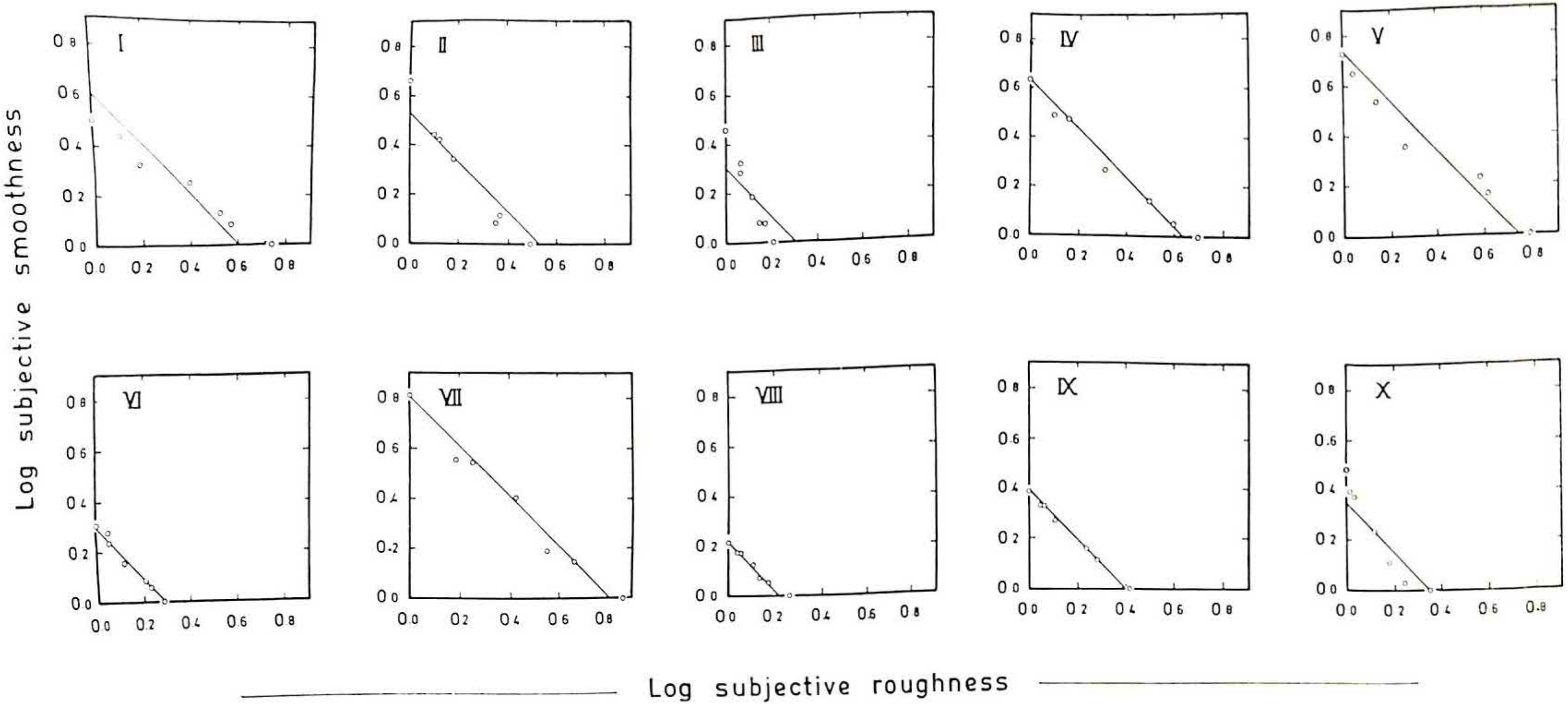


FIG. 3. Subjective smoothness plotted against subjective roughness, in log-log coordinates. (Each graph represents an individual *S*.)

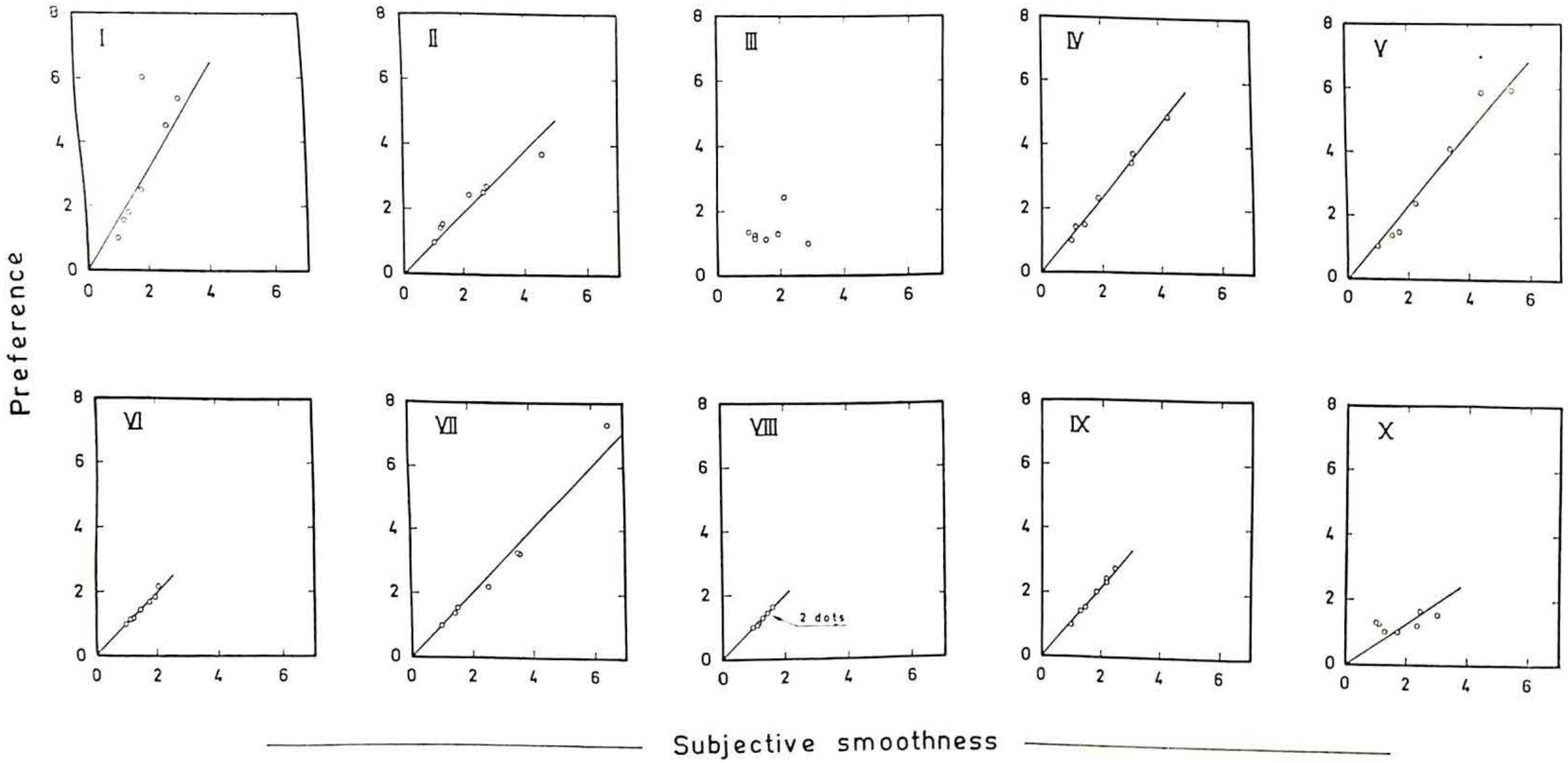


FIG. 4. Preference plotted against subjective smoothness, in linear coordinates. (Each graph represents an individual *S*.)

concerning saltness and sweetness (Ekman & Aakesson, 1964).

Whereas the individual power function may be considered as approximately verified in the present experiment, it is seen from the graphs in Fig. 1 that there is a considerable variation with regard to the slope of the straight lines fitted to the data. In other words, the exponent of the power function is a variable over individuals. In the present sample of 10 *S*s, the exponent of the functions fitted to all seven stimuli varies between .8 and 3.5 for roughness and between −.8 and −3.0 for smoothness. Although part of this variation may be due to unreliability of data, it can be concluded from inspection of the graphs in Fig. 1 and, particularly, Fig. 2 that other sources of variation are much more important.

Reciprocality of scales.—It has been shown in previous investigations by Torgerson (1960, 1961), Eisler (1962), Stevens and Harris (1962), and Schneider and Lane (1963) that magnitude scales tend to be inversely related for continua defined by opposite terms (e.g., loudness and softness). In some cases the inverse relationship is not obtained in reasonable approximation, unless a scale correction is performed (Eisler, 1962; cf., Schneider & Lane, 1963).

Let roughness be denoted x and smoothness y. If the relation is one of simple reciprocality, then $y = c/x$, or $\log y = \log c - \log x$. This hypothesis has been tested by plotting smoothness against roughness, in log-log coordinates, in Fig. 3 where each *S* is represented by a separate graph.

It is seen that for most *S*s the relation, $\log y = \log c - \log x$, describes the data in fairly good approximation; the straight lines drawn in the graphs all have a slope of −1. It may thus be concluded that scales of roughness and smoothness, when constructed by the method of free ratio estimation used in this experiment, tend to be inversely related for individual *S*s.

The preference function.—One of the main purposes of the present investigation was to study preference as a function of the subjective variables of roughness and smoothness. In Fig. 4 the preference variable is plotted against smoothness, in linear coordinates. It is immediately seen that the relation is one of direct proportionality.

Letting p denote preference, the relations between the three subjective

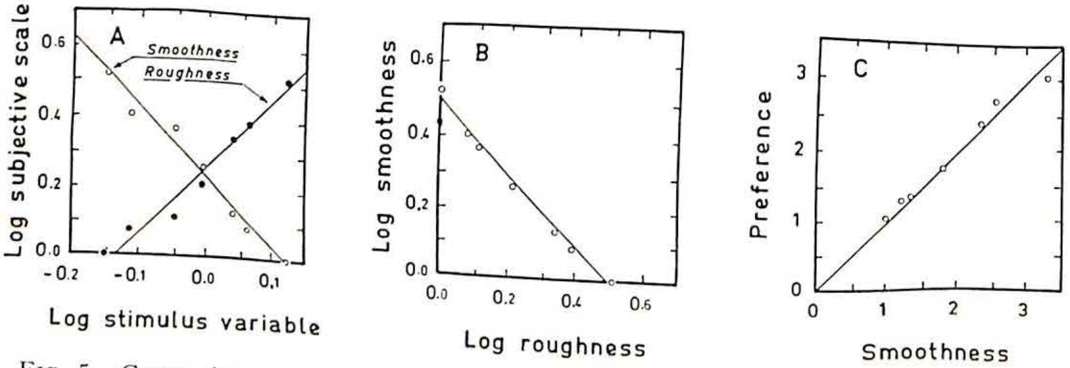


FIG. 5. Group data summarizing results. (Diagram A shows subjective roughness, filled circles, and subjective smoothness, empty circles, plotted against the stimulus variable, in log-log coordinates. Diagram B illustrates, in log-log coordinates, the reciprocality of roughness and smoothness. Diagram C shows preference as a function of smoothness, in linear coordinates.)

variables can now be summarized thus: $p = by = a/x$, where a and b are constants related to the arbitrary units of measurement.

Group data.—Group scales have been obtained by computing geometric means of individual scale values. The psychophysical function, the reciprocality of smoothness and roughness, and the preference function are illustrated by group data in Fig. 5 A–C.

These graphs do not add anything to the results we have already obtained and illustrated. Insofar as we are interested in the generality—over individuals—of our results, the group data convey less information. They are presented here only as a summary of results.

DISCUSSION

The conclusions of the present study have all been based on individual data. This was possible partly because of the reasonably satisfactory regularity of these data. It should be borne in mind that each individual scale was constructed from a ratio matrix based on 4 × 21 or altogether 84 estimates, obtained in the course of one session, lasting about 1 hr. The complete method of ratio estimation—involving all different combinations of stimuli—appears efficient for the construction of reasonably reliable individual scales.

The considerable interindividual variation with regard to the exponent of the power function for roughness and smoothness is reflected also in the scales of preference. Basically it is a variation with regard to the *subjective range* corresponding to the constant stimulus range. A given set of stimuli elicits subjective responses which cover a wide range in some *S*s (e.g., I, V, and VII) and a narrow range in some other *S*s (e.g., III, VI, and VIII); these differences are conspicuous in Fig. 1–4. As far as the interpretation of these findings is concerned, there appear to exist two main alternatives: (*a*) The differences are "perceptually real," i.e., the differences reflect some kind of variation with regard to sensitivity. (*b*) The differences reflect the different ways in which *S*s handle the numbers, i.e., the differences represent systematic errors in the scale construction, basically an inability of some or all of the *S*s to give adequate quantitative reports.

It appears impossible to decide, on the basis of present knowledge, between these alternative interpretations of the interindividual differences which have been found. Similar alternative interpretations have been suggested in an interesting paper by Jones and Marcus (1961), reporting a study of individual power functions for three continua. Positive correlations between exponents were considered as supporting the view that interindividual differences largely reflect differences in handling numbers. On the other hand, the present authors and their co-workers, in unpublished studies of a larger number of continua, have found correlations which are low or moderate as compared to remarkably high reliability coefficients characteristic of individual scales. A further clarification of the issue is highly desirable. Generally speaking, the problem of individual number behavior is basic to vast areas of current psychophysics (cf., Ekman, 1964a, 1964b).

Finally, it is worthy of notice that some of the generalities revealed by the present study may be valid results regardless of the answer to the questions just discussed. The power function might be the true individual psychophysical law, even if the exponents should not be taken at their face values. The reciprocality of smoothness and roughness might be a genuine result, which it was possible to demonstrate because both scales were distorted in the same way. For similar reasons, the simple relation found between the preference variable and the perceptual variables of smoothness and roughness might probably be accepted as a valid result.

REFERENCES

Eisler, H. Similarity in the continuum of heaviness with some methodological and theoretical considerations. *Scand. J. Psychol.*, 1960, **1**, 69–81.

Eisler, H. Empirical test of a model relating magnitude and category scales. *Scand. J. Psychol.*, 1962, **3**, 88–96.

Eisler, H., & Ekman, G. A mechanism of subjective similarity. *Acta psychol., Amsterdam*, 1959, **16**, 1–10.

Ekman, G. Two generalized ratio scaling methods. *J. Psychol.*, 1958, **45**, 287–295.

Ekman, G. A simple method for fitting psychophysical power functions. *J. Psychol.*, 1961, **51**, 343–350. (a)

Ekman, G. Some aspects of psychophysical research. In W. A. Rosenblith (Ed.), *Sensory communication*. New York: Wiley, 1961. Pp. 35–47. (b)

Ekman, G. Is the power law a special case of Fechner's law? *Percept. mot. Skills*, 1964, **19**, 730. (a)

Ekman, G. A. A note on alternative psychophysical laws. Report No. 175, 1964, University of Stockholm, Psychological Laboratories. (b)

Ekman, G., & Aakesson, C. Saltness, sweetness and preference; A study of quantitative relations in individual subjects. Report No. 177, 1964, University of Stockholm, Psychological Laboratories.

Ekman, G., Engen, T., Künnapas, T., & Lindman, R. A quantitative principle of qualitative similarity. *J. exp. Psychol.*, 1964, **68**, 530–536.

Ekman, G., Goude, G., & Waern, Y. Subjective similarity in two perceptual continua. *J. exp. Psychol.*, 1961, **61**, 222–227.

Guilford, J. P. A study in psychodynamics. *Psychometrika*, 1939, **4**, 1–23.

Jones, F. N., & Marcus, M. J. The subject effect in judgments of subjective magnitude. *J. exp. Psychol.*, 1961, **61**, 40–44.

Mashhour, M. *Psychophysical relations in the perception of velocity.* Stockholm: Almqvist & Wiksell, 1964.

Pradhan, P. L., & Hoffman, P. J. Effect of spacing and range of stimuli on magnitude estimation judgments. *J. exp. Psychol.*, 1963, **66**, 533–541.

Schneider, B., & Lane, H. Ratio scales, category scales, and variability in the production of loudness and softness. *J. Acoust. Soc. Amer.*, 1963, **35**, 1953–1961.

Stevens, J. C., & Guirao, M. Individual loudness functions. *J. Acoust. Soc. Amer.*, 1964, **36**, 2210–2213.

Stevens, S. S., & Harris, J. R. The scaling of subjective roughness and smoothness. *J. exp. Psychol.*, 1962, **64**, 489–494.

Torgerson, W. S. Quantitative judgment scales. In H. Gulliksen & S. Messick (Eds.), *Psychological scaling: Theory and applications.* New York: Wiley, 1960. Pp. 21–31.

Torgerson, W. S. Distances and ratios in psychophysical scaling. *Acta psychol., Amsterdam*, 1961, **19**, 201–201.

(Early publication received October 13, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 1, 27–31

RESPONSE FORCE AS A FUNCTION OF AMOUNT OF REINFORCEMENT[1]

VINCENT DI LOLLO

University of Western Australia

W. D. ENSMINGER AND J. M. NOTTERMAN

Princeton University

5 groups of 6 rats each were trained to press an isometric lever and were rewarded with either 1, 2, 3, 4, or 5 20-mg. pellets on responses meeting or exceeding an 8-gm. force criterion. After 10 Acquisition sessions all Ss were shifted to 3 pellets for 6 further sessions and then received 2 Extinction sessions. Both force and variability of bar pressing varied inversely with amount of reward by the end of Acquisition, but not during Shift or Extinction. An account of these results in terms of criterion discrimination is suggested.

The study here reported attempts to determine (*a*) whether force of bar pressing during acquisition varies systematically with magnitude of reinforcement, and (*b*) whether shifts in amount of reinforcement are followed by shifts in force emission.

Exploratory data bearing upon these questions have been reported by Notterman and Mintz (1965) who found an inverse relationship between force and amount of reinforcement. Since in this preliminary study the same *S*s were tested at all levels of amount of reinforcement, the original effects of reward magnitude were conceivably confounded with the effects of shifts in the same variable. In the present research separate groups of *S*s were trained at five independent amounts of reinforcement, were shifted to the middle level, and then placed in extinction.

METHOD

Apparatus.—The apparatus used in this study has been described by Notterman (1959) and by Notterman and Mintz (1965). Essentially, it consists of a Skinner box equipped with a bar connected to a pair of strain gauges. The peak force of each bar press was measured and automatically recorded to the nearest gram. The apparatus was calibrated to record every bar press meeting or exceeding a force of 2.5 gm. (threshold), and to deliver a given number of food pellets following each bar press meeting or exceeding a force of 8.0 gm. (criterion). The experimental room was sound-insulated and was illuminated by a frosted white 7.5-w. indirect lamp located approximately 2 ft. from the experimental cage. Besides the masking noise of an electric fan, the only distinctive sound was the noise of the pellet dispenser which operated upon termination of responses reaching the 8.0-gm. criterion. Water was always available in the experimental cage.

Subjects and procedure.—The *S*s were 36 male Wistar rats obtained from the Charles River Breeding Laboratory. They were 90 days old at the beginning of Acquisition and were housed individually. Fourteen days before the beginning of Acquisition all *S*s were placed on a 24-hr. feeding schedule with free access to dry Purina laboratory mash for 1 hr. each day and with water available at all times. On each of the 2 days immediately preceding the beginning of Acquisition each *S* received approximately 20 min. of tray-approach training. On each approach to the tray *S* obtained the number of pellets it was scheduled to receive during Acquisition.

---

[1] This research was performed under Contract Nonr-1858 (19) between the Office of Naval Research and Princeton University. The first author was in Princeton under the auspices of a Fulbright Travel Grant. Participation of the second author was aided by a National Science Foundation Undergraduate Fellowship.

The 36 *S*s were randomly allocated to six groups of 6 *S*s each. The experiment was carried out over 20 successive days (one session per day) divided into three stages as follows: during *Acquisition* (10 days) Groups I, II, III, IV, and V received either one, two, three, four, or five Noyes 20-mg. pellets, respectively, following each response at or above the 8.0-gm. criterion. As can be seen, number of pellets covaried with the weight of food obtained per response. A control group (Group C) receiving two 45-mg. pellets on responses meeting the 8-gm. criterion was introduced to obtain some indication of the relative effects of weight of food as opposed to the number of pellets. If *S*s respond largely to the weight of the reward, Group C should be expected to behave more like Group V (100 mg.) than Group II (40 mg.).

A "self-shaping" procedure was adopted on the first day of Acquisition: each *S* was placed in the experimental cage where it found the appropriate number of pellets in the food cup, and the bar baited with a small amount of food powder. Subsequent reinforcements depended entirely on the occurrence of responses meeting or exceeding the 8-gm. criterion. A daily session ended upon procurement of 35 reinforcements.

During *Shift* (6 days) all groups received three 20-mg. pellets following responses meeting or exceeding the criterion. The number of reinforcements per session remained unchanged. During *Extinction* (2 days) no food reward was delivered, but the sound of the magazine was present as during Shift. An extinction session ended upon emission of 35 responses meeting or exceeding the 8-gm. criterion. Except for Extinction days, each run began with the relevant number of pellets in the food cup. One *S* in Group III died on the fifth day of shift, and one *S* in Group I was eliminated because of apparatus failure.

RESULTS AND DISCUSSION

The performance of the control group receiving two 45-mg. pellets was entirely similar to that of the five-pellet (100-mg.) group throughout the experiment. Peak force values on the last days of each of the phases were as follows for Group V and Group C, respectively: Acquisition, 10.2 vs. 9.3; Shift, 10.5 vs. 10.7; Extinction, 27.4 vs. 30.1 gm. Following the demonstration that amount of reward, rather than number of pellets, was the relevant independent variable, the results of the control group were excluded from further consideration.

The mean peak force of the responses emitted by *S*s in each group during the last 2 days of Aquisition is shown in the upper set of points (open circles) of Fig. 1.

The two-tailed probability of obtaining an ordered replication of the five reinforcement magnitudes in the force function is $2/5!$ ($p < .02$). This inverse relationship does not appear to be consonant with the direct relationship obtained between amount of reinforcement and either *rate* of bar pressing (Hutt, 1954; Smith & Kinney, 1956) or speed of response in the runway (Crespi, 1942; Di Lollo, 1964), although it is quite comparable to the force data previously reported (Notterman & Mintz, 1965).

Since the present data were obtained under conditions involving an 8-gm. criterion, evidence of effective performance may be sought in a measure of *S*s' approximation of this criterion. Figure 2 shows the distribution of responses for Groups I and V on the last day of Acquisition, and illustrates the greater success of Group V in criterion approximation ($p < .05$, Kolmogorov-Smirnov two-sample test). The variability of the other groups fell between these two extremes as shown in the lower set of points (filled circles) of Fig. 1.

On the basis of these results it may be suggested that the prime effect of amount of reinforcement is to influence *S*s' precision of force emission. The hypothesis that amount of reinforcement influences *S*s' ability to discriminate the criterion is in agreement with other results showing a positive relationship between incen-

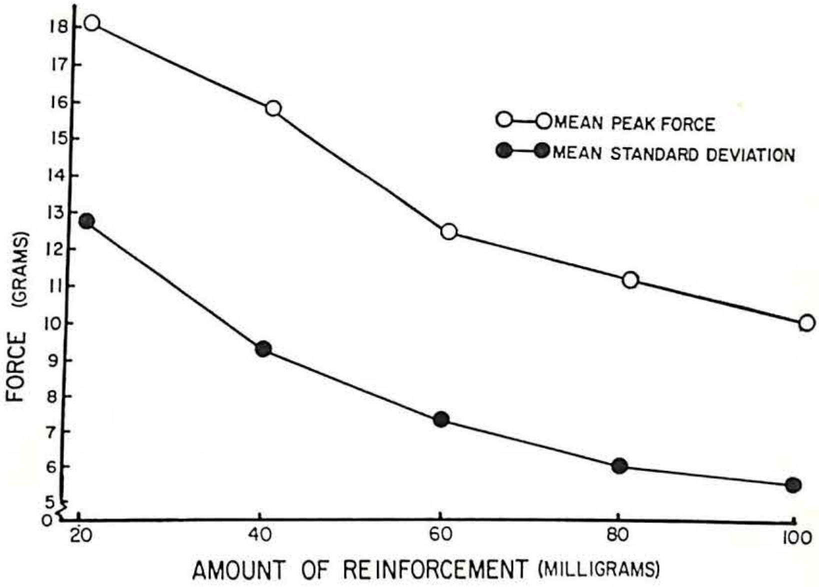


FIG. 1. Mean peak force and *SD* for each group averaged over the last 2 days of Acquisition.

tive amount and effective discrimination in the T maze (Reynolds, 1950; Smith & Duffy, 1957).

It should be noted that the total amount of food obtained during an experimental session varied systematically across groups: *S*s in Group V received 3.5 gm. of food vs. .7 gm. for *S*s in Group I. This could have resulted in long-term differences in the level of deprivation and hence in systematic differences in motivation. Analogous data are available to indicate that substantial differences in hours of food deprivation prior to daily runs influence the force of response, and in a direction appropriate to this argument (Notterman & Mintz, 1965). It may also be reasoned that the effect of amount of reinforcement is principally manifested in "energizing" the animals. Thus, *S*s receiving lesser amounts of food per response are "energized" more than *S*s receiving greater amounts. The emission of forces by the former occasionally reaches higher levels, and consequently variance increases. However, force studies which might support any of these interpretations to the exclusion of others have not been done.

During Shift neither mean peak force nor variability of response differed significantly between groups. There was, however, an overall decrement in the magnitude of both measures over the 6 days of the Shift stage. Table 1 is a summary of the analyses of variance performed on the slope of the linear component of mean peak force and standard deviation, respectively, for each *S* over the 6 days of the Shift stage. Immediately following the reinforcement shift, Groups I and II showed a rapid decrement in the variability of the distribution of forces, with a clustering of responses about the criterion resembling the distributions of the higher-reward groups. Mean peak force decreased accordingly. Groups IV and V showed an increment in the variability of responses—

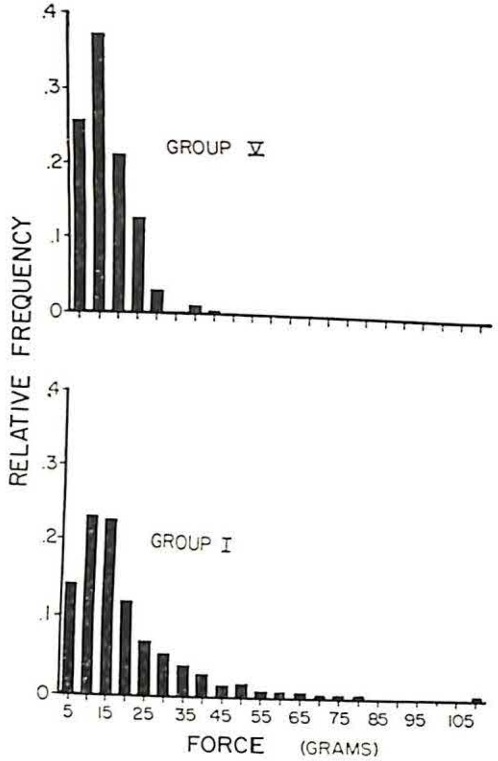


FIG. 2. Force distributions of all responses emitted by Groups I and V on the last day of Acquisition.

and in mean peak force—only at the beginning of the Shift stage, but both measures soon returned to the preshift level. Similar findings have been reported by Notterman and Mintz (1965), who obtained permanent decrements in peak force following increments in reinforcement, but only a temporary increment in force following a reinforcement decrement. This pattern of results is not inconsistent with the discrimination hypothesis proposed in the foregoing: while a reinforcement increment leads to improved discrimination between criterion and noncriterion forces, a reinforcement decrement results in only a temporary disturbance of the established discrimination.

TABLE 1

F Ratios for Slopes of Mean Peak Forces (F_p) and *SD*s over the 6 Days of Shift

Source	*df*	F_p	*SD*
M	1	5.64*	15.30***
Between groups	4	.10	.33
Error (*MS*)	23	(22.7)	(4.6)

* $p < .05$.
*** $p < .001$.

Variability of response force during Extinction is depicted in Fig. 3; mean peak force followed roughly the same trends. Notable is the fact that Group III, which experienced no shift in amount of reinforcement, shows the greatest variability. Perhaps the explanation of this finding lies in the fact that all groups with the exception of Group III have had the cutaneous and kinesthetic stimulus correlates of criterion responding associated, by the onset of Extinction, with two values of amount of rein-

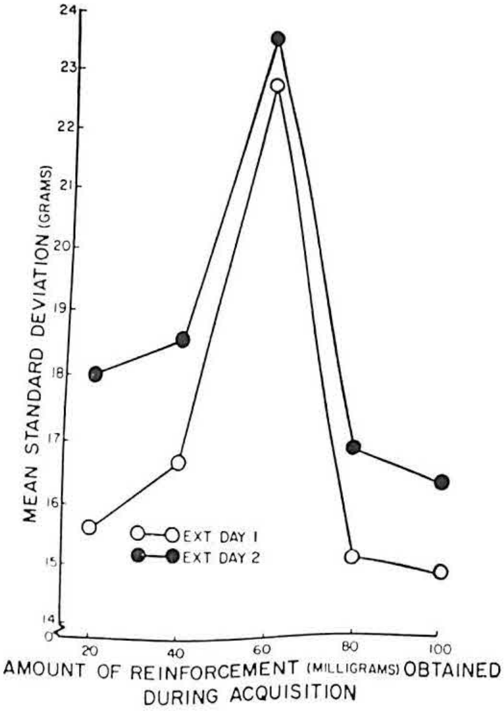


FIG. 3. Mean *SD* for each group during Extinction.

forcement. Therefore, these animals have had opportunity to learn that neither increments (Groups IV and V) nor decrements (Groups I and II) in above-criterion force are related to amount of reinforcement received. The discriminative consequences of the Shift procedure are carried over into the Extinction phase, and are reflected in the lower variability characteristic of Groups I, II, IV, and V. Evidence that it is these groups which are lower than might be expected, and not Group III which is higher, is available in a study describing the behavior of a group of rats placed in extinction following 20 days of CRF (45 mg. reinforcement) at an 8-gm. criterion (Notterman & Mintz, 1965). The mean standard deviation in the cited experiment is 24 gm. on the first 2 days of extinction, almost identical to the values of Group III in the present research.

REFERENCES

Crespi, L. P. Quantitative variation of incentive and performance in the white rat. *Amer. J. Psychol.*, 1942, **55**, 467–517.

Di Lollo, V. Runway performance in relation to runway-goal-box similarity and changes in incentive amount. *J. comp. physiol. Psychol.*, 1964, **58**, 327–329.

Hutt, P. J. Rate of bar pressing as a function of quality and quantity of food reward. *J. comp. physiol. Psychol.*, 1954, **47**, 235–239.

Notterman, J. M. Force emission during bar pressing. *J. exp. Psychol.*, 1959, **58**, 341–347.

Notterman, J. M., & Mintz, D. E. *Dynamics of response.* New York: Wiley, 1965, in press.

Reynolds, B. Acquisition of a simple spatial discrimination as a function of the amount of reinforcement. *J. exp. Psychol.*, 1950, **40**, 152–160.

Smith, M., & Duffy, M. Evidence for dual reinforcing effect of sugar. *J. comp. physiol. Psychol.*, 1957, **50**, 242–247.

Smith, M., & Kinney, G. C. Sugar as a reward for hungry and nonhungry rats. *J. exp. Psychol.*, 1956, **51**, 348–352.

(Early publication received
December 29, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 1, 32–42

HEART RATE DURING CONDITIONING IN HUMANS: EFFECTS OF UCS INTENSITY, VAGAL BLOCKADE, AND ADRENERGIC BLOCK OF VASOMOTOR ACTIVITY[1]

PAUL A. OBRIST, DONALD M. WOOD, AND MARIO PEREZ-REYES

School of Medicine, University of North Carolina

2 experiments concerned with the influence of the vagal and sympathetic innervations of the heart and of peripheral reflex mechanisms are reported on the deceleration of heart rate observed on test trials during classical conditioning in human Ss. An increase in vagal restraint was observed to be the basis of the deceleratory response, which in turn masks the manifestation of sympathetic acceleratory effects. The latter were only observed when the vagus was pharmacologically blocked and when a very intense UCS was used. When the conditioned pressor responses were blocked pharmacologically, as evaluated by direct recordings of arterial blood pressure, the deceleratory response was not changed, indicating that the peripheral homeostatic reflex mechanisms are not the basis for this response and that it is likely a conditioned response.

The physiological basis of heart-rate changes in human *S*s during classical conditioning was investigated in order to determine the influence of the vagal and sympathetic innervations of the heart and of peripherally initiated homeostatic reflexes. The heart-rate changes regarded as conditioned, i.e., those changes occurring either between CS and UCS onset or for a more sustained period following CS onset on nonreinforced trials, are in certain respects paradoxical and subject to the question of whether they are conditioned responses or only artifacts of the conditioning process. When a noxious UCS is used, such responses are reported to be either a deceleration of heart rate or a biphasic response characterized by an initial acceleration followed by a usually more sustained deceleration (Deane, 1964; Fuhrer, 1964; Geer, 1964; Lang & Hnatiow, 1962; Notterman, Schoenfeld, & Bersh, 1952; Wilson, 1964; Wood & Obrist, 1964; Zeaman, Deane, & Wegner, 1954; Zeaman & Smith, 1964; Zeaman & Wegner, 1957). Recently, Wood and Obrist (1964) have demonstrated that the initial acceleratory component of the biphasic response is a respiratory artifact, which leaves only the deceleration as a possible conditioned response.

Two questions now arise. First, why is there no sympathetic acceleratory response since, under these conditions, such a sympathetic effect would be expected and was observed as measured by GSR and vasomotor responses? There is evidence that vagal inhibition, manifested as a deceleration of heart rate, is able to mask sympathetic activity (Rushmer, 1958; Samaan, 1934–35a). Therefore, the purpose of the first of two

---

[1] These studies were supported principally by Research Grants M-6020-A and MH-07995, National Institutes of Health, United States Public Health Service, and by faculty grants from the University of North Carolina. A portion of this paper was read at the Society for Psychophysiological Research meetings, October 1964. The technical assistance of James Howard and Norma Thomas is gratefully acknowledged.

experiments to be reported was to determine whether sympathetic effects are present during conditioning but are not manifested due to the dominance of vagal restraint. For this purpose, heart-rate changes during conditioning in *S*s with an intact vagus are compared to those with vagal activity pharmacologically blocked. Additionally, two intensities of the UCS were used to determine whether sympathetic effects become more pronounced with a far more intense UCS than was previously used (Wood & Obrist, 1964). It was anticipated that a more intense UCS would result in an observable sympathetic effect even with the vagus intact, due to the possibility of greater fear or anxiety-producing qualities of the stimulus (Martin, 1961). The conditioned GSR was also evaluated to demonstrate the presence of sympathetic activity.

A second question concerned whether the deceleratory response is a conditioned response or rather is artifactual, resulting from homeostatic reflex mechanisms initiated by a conditioned pressor response. The use of a noxious UCS as well as the evidence of conditioned vasomotor responses indicate that such pressor responses may occur. Further, the presence of such reflex influences on heart rate is well documented (Neil & Heymans, 1962), but their significance during conditioning has not been determined. Therefore, the second experiment reported was to determine whether the deceleratory response was still observed once pressor responses of vasomotor origin were pharmacologically blocked. In order to evaluate the effectiveness of the pharmacological agent, direct recordings of both diastolic and systolic blood pressure were obtained.

EXPERIMENT I

Method

Subjects.—The *S*s were 60 healthy male undergraduate students randomly selected from the introductory psychology class.

Apparatus.—Methods for measuring heart rate, respiration, and skin resistance have been previously described (Obrist, Hallman, & Wood, 1964; Wood & Obrist, 1964). Additionally, R-R interval time was measured to the nearest .001 sec. on a sequence event timer and recorder (SETAR; Welford, 1952) and recorded on punch tape for computer analysis.

Procedure.—There were four experimental conditions with *S*s assigned randomly to each. This provided a comparison of the effects of vagal and nonvagal blockade at two intensities of the UCS. Vagal blockade was achieved by an IV injection of atropine sulfate (.02 mg/kg of body weight) just prior to the first conditioning trial. The dosage was derived from pilot studies. The *S*s not given atropine received IV 5 cc of saline but were told that it was atropine. Eleven *S*s were eliminated from the experiment prior to data analysis either because they could not relate correctly which CS was followed by shock, or the atropine in part or whole was administered subdermally. Therefore, the final N was: saline, less intense UCS, $N = 12$; vagal block, less intense UCS, $N = 11$; saline, more intense UCS, $N = 13$; and vagal block, more intense UCS, $N = 13$.

1. UCS intensity. The intensity of the UCS was adjusted for each *S* as follows. All *S*s were given a 9-point scale extending from "just felt" to "very painful" by which to judge a series of shocks. They were then given an ascending series of 6-sec., 60-cycle shocks spaced 30 sec. apart and administered to the finger tips of the left hand through two dry 5-mm. diameter silver EEG electrodes. At random intervals, the intensity was decreased one step to determine the reliability of *S*'s judgments. At the point where the shock was first judged as "very painful," the series was stopped. For *S*s in the low-intensity UCS groups, this was the level to be used during conditioning and *S* was so informed. This intensity approximated the level used in the Wood and Obrist (1964) study. For *S*s in the high-intensity UCS groups, *E* entered the *S* room following the "very painful" judgment and read an appeal for *S* to take considerably more shock, which was defined to *S* as "as much as you can humanly stand." Thus, an appeal was made to his masculinity in addition to offering him a substantial financial bonus. After the first

five conditioning trials, these *Ss* were again asked to take one further increase (.3 ma.), but most refused. The average amount of current for the less and more intense conditions was 1.88 and 3.86 ma., respectively ($t = 10.61$, $p < .001$).

2. Respiratory control. Respiration was controlled throughout conditioning by requiring *S* to maintain his normal resting frequency and depth so as to minimize the influence of respiratory effects on heart-rate changes in *Ss* not receiving vagal block. These procedures are similar to those previously reported (Wood & Obrist, 1964), with one exception. Respiratory activity was paced by an auditory signal which mimicked human breathing sounds. This signal could be adjusted with respect to both inspiration and expiration time as well as the pause between each so as to follow exactly any one *S*'s resting pattern.

3. Conditioning procedures. A 10-min. rest period, training for respiratory control, UCS intensity determination and a 10-trial adaptation series preceded, in that order, differential trace conditioning procedures. During adaptation each CS was presented five times in a fixed random order. Conditioning consisted of a series of 36 trials, with 16 reinforced, 10 test, and 10 sensitization control trials, presented in a fixed random order. Short rest periods were given after every 12 trials. Test trials were used so as to determine the heart-rate changes during the period when the UCS would normally occur, since the maximum deceleration occurs during this time (Wood & Obrist, 1964). The CS+ and CS− were dim red and blue lights with a duration of 2 sec. The 6-sec. UCS started 7 sec. after CS onset. The intertrial interval averaged 75 sec. with a range of 60 to 90 sec. These intervals were varied in a fixed random order. Except during instructions, a low-level white masking noise was used. Respiratory control was continuous except during rest periods. The *Ss* were instructed that only one of the lights would be followed by shock and then only some of the time.

Data quantification.—The following are data quantification procedures.

1. Heart rate. Second-by-second changes in R-R interval from a pre-CS base line for each of the 25 sec., following CS onset were averaged separately for the 5 CS+ and 5 CS− adaptation trials, 10 test, 10 sensitization control, and 6 of the UCS trials, using procedures previously described (Wood & Obrist, 1964). This period was used in order to observe long latency sympathetic effects (see Dykman & Gantt, 1959). These analyses were done for each experimental condition except adaptation. Here the saline and vagal blockade groups were combined for each level of UCS intensity since atropine had not yet been administered. This increased the *N* to 23 in one group and 25 in the other. As in the previous study (Wood & Obrist, 1964), heart-rate changes, either acceleratory or deceleratory, were not related to base level. Therefore, no correction for base level was used.

2. Respiration. A mean change score was calculated as a ratio of the pre-CS average respiratory amplitude to the single largest inspiration during the first 6, next 9, and last 10 sec. following CS onset. This was calculated separately for test, sensitization, and UCS trials. This method of quantification is more completely described elsewhere (Wood & Obrist, 1964).

3. GSR. Conditioned GSR activity was quantified for test and sensitization control trials by first obtaining the highest resistance level (ohms) up through 2.0 sec. following CS onset and the lowest resistance level occurring between 2 and 10 sec. after CS onset. In cases where there were two or more responses having a 600-ohm or greater change and a recovery of at least 600 ohms, the lowest level was obtained by totaling the magnitude of all such responses and subtracting this from the highest level. These two levels were then averaged for each *S* separately for test and sensitization control trials and then converted to conductance units. The difference between these two-level scores was then obtained as the measure of GSR magnitude and then averaged for only the two experimental groups not receiving the vagal blockade, since atropine blocks GSR activity. This resulted in an *N* too small to correct for base-level effects by regression analysis (Lacey & Lacey, 1962).

Results

No significant differences were observed in the second-by-second heart-rate changes between the CS+ and CS− during adaptation. Therefore, the sensitization control (CS−) was used as a base line from which heart-rate changes on nonreinforced test trials and UCS trials could be evaluated. A direct comparison of the heart-rate changes between vagal blockade and nonvagal blockade groups was prohibited by the very large difference in variance between groups. Base-level heart rate is greatly elevated by vagal blockade, resulting in a reduction of responsiveness. However, since the principal effects of vagal blockade were expected to be a shift in the direction of the response, that is, from deceleration to acceleration, an evaluation of the reliability of heart-rate changes within each experimental group makes direct comparisons unnecessary.

The deceleration of heart rate during conditioning was found to be the result of an increase in vagal restraint which masked sympathetic effects. This result was most obvious with the more intense UCS condition. The high-intensity UCS group without vagal blockade showed a biphasic response. There was an initial acceleration extending from Sec. 2–6, the effect being significant on Sec. 3 and 4. This was followed by a more sustained deceleration extending from Sec. 7–15, the effect being significant on Sec. 8 and 9. On the other hand, when the vagus was blocked, only an acceleratory response was observed on test trials with the exception of the first second (see Fig. 1 and Table 1). The acceleration was significant on Sec. 4–10, peaking at the exact second that the deceleration peaked when the vagus was not blocked, i.e., Sec. 9. There was a second less pronounced acceleration which began around Sec. 16 and peaked on Sec. 21, the ampli-

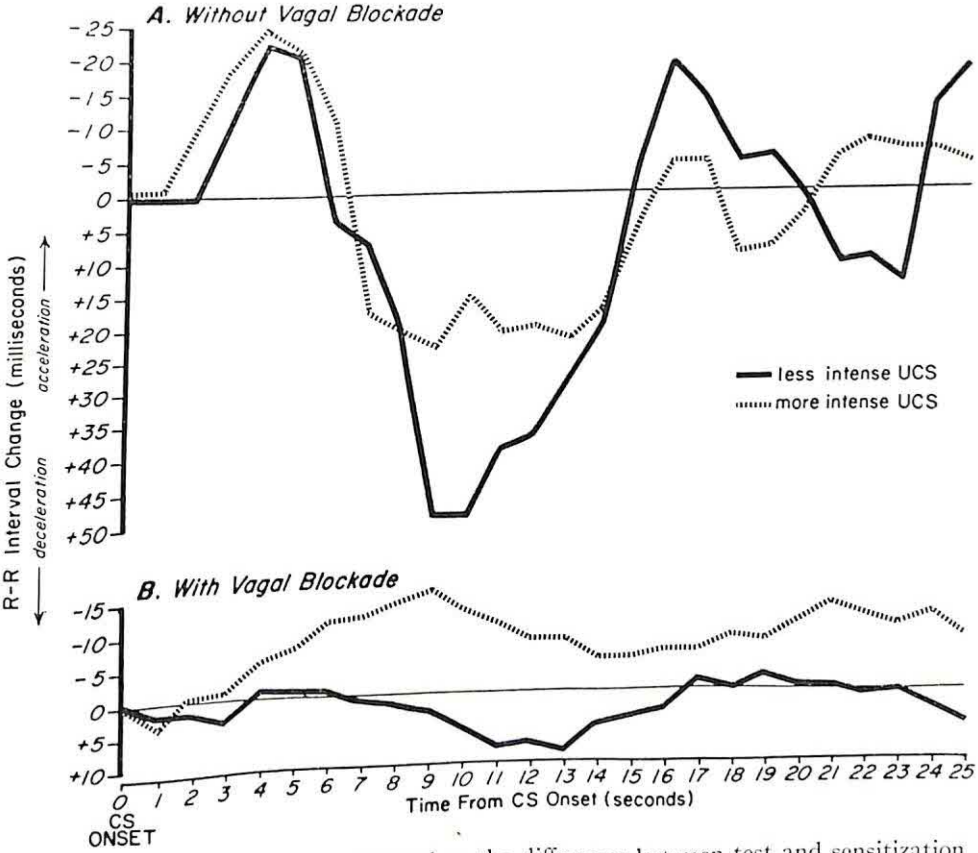


FIG. 1. Heart-rate changes expressed as the difference between test and sensitization control trials with and without vagal blockade at two intensities of the UCS.

TABLE 1

SECOND BY SECOND CHANGES IN R-R INTERVAL (MILLISECONDS) EXPRESSED AS TEST MINUS SENSITIZATION CONTROL TRIAL MEAN DIFFERENCES

Sec. since CS Onset	Low-Intensity UCS				High-Intensity UCS			
	No Vagal Blockade $N = 12$		Vagal Blockade $N = 11$		No Vagal Blockade $N = 13$		Vagal Blockade $N = 13$	
	$\bar{x}$	t	$\bar{x}$	t	$\bar{x}$	t	$\bar{x}$	t
1	0	<1	+2	1.17	0	<1	+4	3.49**
2	0	<1	+2	<1	−10	1.34	−1	<1
3	−11	<1	+4	1.07	−19	2.79*	−1	<1
4	−22	1.47	−1	<1	−24	2.54*	−5	2.30*
5	−20	1.25	−1	<1	−21	1.70	−7	2.45*
6	+4	<1	−1	<1	−11	1.25	−11	2.69*
7	+8	<1	+1	<1	+17	1.76	−11	2.89*
8	+21	1.46	+2	<1	+21	2.50*	−13	2.53*
9	+50	4.01**	+3	<1	+23	2.78*	−15	2.85*
10	+50	3.20**	+6	1.88	+16	1.38	−12	2.40*
11	+39	2.87*	+9	2.72*	+21	1.60	−11	2.00
12	+37	2.94*	+8	2.28*	+20	1.51	−8	1.66
13	+28	2.31*	+10	3.55**	+22	1.55	−8	1.92
14	+20	1.70	+6	2.58*	+18	1.15	−5	1.32
15	−5	<1	+5	2.15	+5	<1	−5	1.41
16	−20	1.93	+4	1.92	−4	<1	−6	1.50
17	−14	1.34	−1	<1	−5	<1	−5	1.30
18	−4	<1	+1	<1	+10	1.12	−8	1.93
19	−5	<1	−2	<1	+9	1.41	−7	1.62
20	+1	<1	0	<1	+4	<1	−10	2.02
21	+12	<1	0	<1	−5	<1	−12	2.40*
22	+11	<1	+1	<1	−7	<1	−11	2.18*
23	+14	<1	+1	<1	−5	<1	−9	1.62
24	−13	1.16	+3	<1	−6	<1	−11	1.83
25	−19	1.23	+6	1.23	−4	<1	−8	1.33

Note.—+ = longer R-R interval or heart-rate deceleration; − = shorter R-R interval or heart-rate acceleration.
* $p < .05$.
** $p < .01$.

tude being significant on both Sec. 21 and 22.

The acceleration on Sec. 2–6 after CS onset observed with the high-intensity UCS with the vagus intact appears to be a momentary decrease in vagal restraint rather than a sympathetic effect. First, it has a shorter latency and peaks considerably sooner than the sympathetic effect observed with the vagus blocked. Such a latency difference has been previously proposed by Samaan (1934–35b) and Dykman and Gantt (1959). Second, this acceleration appears to be significantly influenced by uncontrolled respiratory effects. The amplitude of the acceleration is positively correlated (rho = .62, $p < .05$) with the largest single inspiration during the period of acceleration, the latter showing an 8% increase ($t = 2.02$, $p < .10$). In contrast, during vagal blockade, there was no correlation between the magnitude of the acceleration and respiratory amplitude. Therefore, even with a very intense UCS, there is little evidence that sympathetic effects are manifested when the vagal innervation is intact.

A reliable deceleratory response was also found on test trials with the less intense UCS without vagal blockade, extending from Sec. 6–14 (see Fig. 1

and Table 1) and peaking at Sec. 9, as in the higher UCS intensity condition. In contrast, little responsiveness was observed in the lower intensity UCS condition with vagal blockade. A small deceleration, significant on Sec. 11–14, suggests that there is little sympathetic involvement at this UCS intensity, at least with respect to the heart. This latter deceleration appears due to a reduction in the amount of vagal blockade during the later test trials which allows some vagal inhibition to become manifested. An analysis of heart-rate changes was performed using only the first two test and sensitization control trials, when vagal blockade is at its maximum as measured by basal levels of heart rate. This revealed no reliable response of any kind, although the mean second-by-second effects are acceleratory from the sixth second on. This is in contrast to a large deceleratory response, which was already significant on 1 sec. in the low-UCS-intensity group without vagal blockade during the same trials.

The lower and higher UCS intensity conditions without vagal blockade were not reliably differentiated on any of the 25 sec. following CS onset (see Table 2). Only during the deceleratory phase of the response are any appreciable differences observed. But this is only a difference in degree of vagal restraint. On the other hand, when the vagus is blocked the two conditions are sharply differentiated, with the higher intensity condition having a significantly greater acceleratory response on Sec. 7–16 (see Table 2). Therefore, differences in UCS intensity result in different degrees of sympathetic activity, while having a less pronounced effect on vagal activity.

TABLE 2

SECOND-BY-SECOND DIFFERENCE IN R-R INTERVAL (MILLISECONDS) BETWEEN THE LOWER AND HIGHER UCS INTENSITIES DETERMINED FROM TEST MINUS SENSITIZATION CONTROL TRIAL MEAN DIFERENCES

Sec.	Without Vagal Blockade		With Vagal Blockade	
	$\bar{x}$	t	$\bar{x}$	t
1	0	<1	− 2	<1
2	−10	<1	− 3	<1
3	− 7	<1	− 4	1.25
4	− 2	<1	− 4	1.36
5	− 1	<1	− 6	1.68
6	−15	1.07	−10	2.03
7	+ 9	<1	−12	2.49*
8	0	<1	−15	2.32*
9	−27	1.83	−18	2.80**
10	−34	1.80	−18	2.91**
11	−18	<1	−19	3.00**
12	−17	<1	−16	2.65*
13	− 6	<1	−18	3.49**
14	− 2	<1	−10	2.39*
15	+10	<1	− 9	2.22*
16	−16	1.26	− 9	2.09*
17	− 9	<1	− 4	<1
18	−14	<1	− 8	1.74
19	−14	1.56	− 6	1.08
20	+ 3	<1	−10	1.67
21	−16	1.14	−12	1.90
22	−18	1.05	−12	1.77
23	−20	1.35	− 9	1.34
24	− 8	<1	−14	1.83
25	−15	<1	−14	1.77

Note.— − = higher intensity, shorter R-R interval or faster heart rate.
* $p < .05$.
** $p < .01$.

Vagal restraint was also observed to decrease the duration of the unconditioned response. In all conditions, a reliable acceleration of heart rate occurred to the UCS which was significantly greater for the more intense UCS both with and without vagal blockade. This acceleration was significant ($p < .05$ or less) with the vagus intact on 3 sec. for the less intense UCS and on 8 sec. for the more intense UCS. On the other hand, with the vagus blocked the acceleration was significant on 8 and 14 sec. for these respective conditions. Furthermore, reliable differences between the acceleratory responses to the two UCS intensities were observed on only 3 sec. with the vagus intact, but on 12 sec. with the vagus blocked.

Sympathetic activity was manifested at both UCS intensities as measured by GSR activity. A greater response on test than on sensitization control trials was observed in 11 of 12 *S*s with the less intense UCS and in all *S*s with the more intense UCS. There was some evidence of greater responsiveness with the more intense UCS, though no reliable differences could be demonstrated. The average number of measurable responses (i.e., those greater than 600 ohms) on the 10 test trials was 7.8 for the less intense UCS and 9.6 for the more intense UCS. Similarly, the mean response amplitude was greater with with the more intense UCS, but the difference was not reliable (Mann-Whitney $U > .10$). Therefore, the failure of sympathetic effects to be manifested by heart rate with the less intense UCS cannot be attributed to the absence of such effects.

EXPERIMENT II

Method

Subjects.—The *S*s were 11 healthy young-adult male student volunteers preselected on the basis of having shown an appreciable deceleratory response during a previous conditioning session, a condition considered necessary to detect clearly the influence of homeostatic effects.

Equipment.—The same equipment was used as in Exp. I in addition to a Statham 0–75 cm. Hg strain gauge for the direct recording of arterial blood pressure.

Procedure.—Two conditioning sessions were required and were separated by at least 1 wk. The first was used only to preselect *S*s. During the second, pressor responses of vasomotor origin were blocked pharmacologically after the eighth test trial by an IV injection of 0.6 mg. of Hydergine-Sandoz, the maximum recommended clinical dosage. (This agent has alpha-receptor adrenergic blocking characteristics and is a combination of three ergot alkaloids.) Conditioning procedures were similar to those reported in Exp. I. The UCS was set at the less intense level. Thirteen test trials were used. Because pressor responses can also be initiated by respiratory-induced cardiac acceleration, special efforts were made to keep respiratory activity controlled in both conditioning sessions. During the second session, sensitization control trials were omitted for five *S*s prior to drug administration and for eight *S*s following drug administration in order to expedite proceedings necessitated by the experiment. Therefore, heart-rate changes were determined for all *S*s without correction for sensitization effects.

In order to evaluate the effectiveness of the adrenergic blockade, direct readings of arterial blood pressure were attempted in the last eight *S*s and successfully obtained in five. Pressure was measured from the radial artery near the wrist of the right arm. The puncture was done under local anesthesia (Xylocaine) using a No. 18 cordon needle. The needle was attached to a strain guage by a catheter filled with heparin saline solution for flushing.

Data quantifications.—Data were quantified as in the previous study. Both diastolic and systolic blood pressure were quantified on a second by second basis, as with heart rate. Heart rate and blood pressure changes on the first three test trials prior to drug administration were not quantified because *S*s were usually less responsive on these trials. This left five test trials before, and five after drug administration to be quantified.

Results

There was no evidence to indicate that the deceleration of heart rate was the result of reflexes initiated by a pressor response. There was initially observed in all 5 *S*s on whom measurements were obtained a small pressor response which had a mean peak response of 1.9 mm. Hg systolic and 1.2 mm. Hg diastolic (see Table 3). Adrenergic blockade reduced the response in all *S*s, with the mean peak response being 0.3 mm. Hg, systolic and diastolic. On the other hand, a reliably large sustained deceleration of heart rate was observed both before and after adrenergic blockade (see Table 4). Based on all 11 *S*s, there were clearly no reliable differences between the deceleratory re-

TABLE 3

Second-by-Second Mean Blood Pressure Levels (Mm. Hg) on Test Trials before and Following Adrenergic Blockade ($N = 5$)

Sec.	Systolic		Diastolic	
	Preblock	Postblock	Preblock	Postblock
$\bar{x}$ Pre[a]	121.9	114.6	73.8	70.1
1	121.5	114.5	73.7	69.9
2	121.4	114.3	73.7	70.4
3	121.9	114.3	74.5	70.2
4	123.0	114.8	74.9	70.4
5	123.2	114.9	75.0	70.4
6	123.3	114.9	74.9	69.9
7	123.6	114.8	75.0	70.3
8	123.8	114.8	74.3	69.7
9	122.8	114.6	73.4	69.3
10	121.9	112.4	73.1	69.1
11	122.0	112.8	72.6	69.3
12	120.9	113.1	72.4	68.8
13	120.8	113.4	72.2	68.8
14	119.6	113.2	72.1	69.0
15	120.9	112.9	72.7	69.6

[a] $\bar{x}$ Pre = Mean of all readings in the 15 sec. before CS onset.

sponses pre- and postblockade. A comparable effect was found with the 5 *S*s on whom measurements of blood pressure were obtained. However, because of the small *N*, the reliability of this effect was not evaluated.

The independence of pressor responses and heart-rate changes is indicated by still other aspects of the data. First, there were trials even prior to adrenergic blockade when heart rate was observed to decelerate even when no pressor response occurred. Second, in one *S*, the pressor response was completely suppressed following blockade but the deceleratory response was unaltered. Third, no correlation could be found between the amplitude of the heart-rate deceleration and the amplitude of either the systolic or the diastolic response, when these correlations were performed within *S*s using responses on test trials. There was a reduction in the deceleratory response in one *S*. However, this appears attributable to the effects of *S*'s knowing when he received the adrenergic blocking agent (all *S*s had been informed of the pos-

TABLE 4

Second-by-Second Changes in R-R Interval (Milliseconds) on Test Trials before and Following Adrenergic Blockade Expressed as Difference from Pre-CS Base Level ($N = 11$)

Sec.	Preblockade		Postblockade		Difference Post − Pre	
	$\bar{x}$	t	$\bar{x}$	t	$\bar{x}$	t
1	+ 6	<1	+ 7	1.37	+ 1	<1
2	+12	<1	− 4	<1	−16	<1
3	−10	<1	+ 3	<1	+13	1.20
4	+ 2	<1	+16	<1	+14	<1
5	+ 9	<1	+ 9	<1	0	<1
6	+12	1.08	+19	1.31	+ 7	<1
7	+33	3.30**	+40	2.39*	+ 7	<1
8	+41	4.24**	+59	4.00**	+18	<1
9	+45	4.43**	+65	4.84**	+20	1.24
10	+52	4.95**	+46	3.83**	− 6	<1
11	+47	4.08**	+42	2.81*	− 5	<1
12	+60	4.29**	+49	4.63**	−11	1.09
13	+43	4.66**	+38	3.91**	− 5	<1
14	+31	3.11**	+26	2.17*	− 5	<1
15	+29	3.03**	+24	2.37*	− 5	<1

Note.— + =longer R-R interval or heart-rate deceleration; − =shorter R-R interval or heart-rate acceleration.
* $p < .05$.
** $p < .01$.

sible postexperimental side effects of this drug, which can include hypotensively induced nausea and fainting). In this *S*, a similar reduction in the deceleratory response was observed to a saline injection on a third conditioning session. However, no reduction in the response was observed to the adrenergic blocking agent when it was then administered following the saline and when the deceleration had returned to its original amplitude.

DISCUSSION

These data indicate that the heart-rate changes observed on test trials during classical conditioning under these experimental conditions are primarily manifestations of variations in vagal activity and that sympathetic effects are only clearly manifested when a very intense UCS is used and then only when the vagal innervation is blocked. Further, the more dominant response is an increase in vagal restraint (i.e., heart-rate deceleration), especially with a less intense UCS, and this response appears to be conditioned. Considering the decelerative response as indicative of, or mediated by, some affective response such as "experimental anxiety" (Deane, 1964; Notterman et al., 1952) is contraindicated by the fact that this is a vagal response. Traditionally, sympathetic activity has been at least implicitly assumed to be involved in such affective responses. This assumption is supported by the greater sympathetic involvement in the high-shock condition when the vagus is blocked, a condition which should result in a far greater affective response. On the other hand, the deceleratory response is greater under the low-shock condition, or when the least affective response might be found.

There are three other ways the decelerative heart-rate change might be viewed. First, it reflects the conditioning of homeostatic processes and as such is a type of exteroceptive-interoceptive conditioning. The UCS in this case is the pressor response to the shock while the UCS would involve CNS activity which slows heart rate. The activity of the latter is only manifested on test trials when respiratory and other effects of the UCS are minimal. This would mean that in the absence of pressor responses, as following adrenergic blockade, the CS is able to initiate anticipatory restraining activity. That such a peripheral response might be an effective UCS is suggested by data reported by Bykov (1957) where pharmacological agents were used to initiate the response. It is also reported that conditioning of this type requires from 20 to 100 reinforcements. However, in the present study, a reliable deceleratory response was observed after only four reinforcements, a fact which argues against this position.

A second explanation has been suggested by Zeaman and Smith (1964), who have observed heart-rate deceleration in anticipation of shock and both pleasant and unpleasant auditory stimuli, using a variation on standard conditioning procedures. It is proposed that this deceleration is mediated first by an attention process initiated by the CS which then initiates respiratory changes resulting in a heart-rate deceleration. However, the studies reported by Zeaman and Smith do not appear to be definitive with respect to the influence of respiration and are contrary to evidence from the Wood and Obrist (1964) study, as well as data recently reported by Deane (1964). Data from the present experiments also contraindicate any influence of respiration on the decelerative response. For example, in both experiments respiration amplitude was uncorrelated with the decelerative response and, with the exception of the more intense UCS conditions of Exp. I, remained perfectly controlled. Therefore, respiration does not appear to mediate whatever influence attention might have on cardiovascular processes.

A third possibility is suggested by some recent work of Lacey (Lacey, 1959; Lacey, Kagan, Lacey, & Moss, 1963; Lacey & Lacey, 1958) where it has been proposed that cardiovascular activity influences environmental interaction via visceral afferent control of central activity. In part, this hypothesis is based on the observation that various exteroceptive stimuli, some with clear affective value, result in both sympathetic discharge, as measured galvanically, and in cardiac deceleration (Davis & Buchwald, 1957; Lacey et al., 1963; Obrist, 1963). This deceleration of heart rate is thought to facilitate sensory intake. Therefore, the deceleration observed during conditioning could involve similar processes and have a like function. As such, it would be a conditioned response in anticipation of sensory intake (i.e., the UCS). However, there is one aspect of the conditioning data which is not consistent with this view and which, for that matter, is yet another unresolved paradox of heart-rate conditioning. This is, the deceleratory conditioned response is in the opposite direction from the UCR, which is always reported as acceleratory, a situation seemingly without parallel in the conditioning of striate and other visceral responses. Therefore, the significance of the deceleratory response is still not understood. Nonetheless, the positions of Lacey and of Zeaman and Smith appear at this time to merit further evaluation. Among other things, the peripheral and central processes involved with the UCR would appear necessary to study.

Recently, Geer (1964) has suggested that the deceleratory heart-rate change observed during conditioning is nothing more than the orienting response to the CS. This position is not supported by the present study. An assessment of such effects, based on the heart-rate response both to the first adaptation trial and to the sensitization control for the less intense UCS, reveals a significant deceleratory response which peaks at the fifth second and which has become a small nonsignificant acceleratory response by the eighth second. In contrast, the deceleratory response on test trials has a longer latency, a larger amplitude, and a longer duration with both intensities of the UCS.

Wood and Obrist (1964) have questioned the suitability of differential conditioning to assess sensitization effects and have suggested that the CS− might be significantly influenced by stimulus generalization from the CS+. Data from the present study also support this possibility. For example, there was no heart-rate change to the CS− during conditioning for the higher intensity UCS saline condition, even though a reliable deceleratory response was observed in the same condition during adaptation and to the CS− during conditioning for the less intense UCS saline condition. Respiration did not appear to be a significant factor. Also, based on the adaptation data, the magnitude of the sensitization effect (i.e., influence of UCS intensity) appears to be directly related to the magnitude of the deceleration. Stimulus generalization appears to be a possible basis for this lack of response to the CS− during conditioning since it would be expected to be greater with the more intense UCS. In part, the failure of the initial acceleratory response to differentiate on test trials between the less and more intense UCS when the vagus was intact could be attributed to the failure to assess sensitization effects.

REFERENCES

Bykov, K. M. *The cerebral cortex and the internal organs.* (Trans. by W. H. Gantt) New York: Chemical Publishing, 1957.

Davis, R. C., & Buchwald, A. M. An exploration of somatic response patterns: Stimulus and sex differences. *J. comp. physiol. Psychol.*, 1957, **50**, 44–52.

Deane, G. E. Human heart rate responses during experimentally induced anxiety: A follow-up with controlled respiration. *J. exp. Psychol.*, 1964, **67**, 193–195.

Dykman, R. A., & Gantt, W. H. The parasympathetic component of unlearned and acquired cardiac responses. *J. comp. physiol. Psychol.*, 1959, **52**, 163–167.

Fuhrer, M. J. Differential verbal conditioning of heart rate with minimization of changes in respiratory rate. *J. comp. physiol. Psychol.*, 1964, **58**, 283–289.

Geer, J. H. Measurement of the conditioned cardiac response. *J. comp. physiol. Psychol.*, 1964, **57**, 426–433.

LACEY, J. I. Psychophysiological approaches to the evaluation of psychotherapeutic process and outcome. In F. A. Rubenstein & M. G. Parloff (Eds.), *Research in psychotherapy*. Washington, D, C.: American Psychological Association, 1959. Pp. 160–208.

LACEY, J. I., KAGAN, J., LACEY, B. C., & MOSS, H. A. Situational determinants and behavioral correlates of autonomic response patterns. In P. J. Knapp (Ed.), *Expression of the emotions in man*. New York: International Univer. Press, 1963. Pp. 161–196.

LACEY, J. I., & LACEY, B. C. The relationship of resting autonomic activity to motor impulsivity. *Res. Publ. Ass. Nerv. Ment. Dis.*, 1958, **36**, 144–209.

LACEY, J. I., & LACEY, B. C. The law of initial value in the longitudinal study of autonomic constitution: Reproducibility of autonomic responses and response patterns over a four year interval. *Ann. N. Y. Acad. Sci.*, 1962, **98**, 1257–1290, 1322–1326.

LANG, P. J., & HNATIOW, M. Stimulus repetition and heart rate response. *J. comp. physiol. Psychol.*, 1962, **55**, 781–785.

MARTIN, B. The assessment of anxiety by physiological-behavioral measures. *Psychol. Bull.*, 1961, **58**, 234–255.

NEIL, E., & HEYMANS, C. Cardiovascular and pulmonary reflexes. In A. A. Luisada (Ed.), *Cardiovascular functions*. New York: McGraw-Hill, 1962. Pp. 103–123.

NOTTERMAN, J., SCHOENFELD, W., & BERSH, P. Conditioned heart rate responses in human beings during experimental anxiety. *J. comp. physiol. Psychol.*, 1952, **45**, 1–8.

OBRIST, P. A. Cardiovascular differentiation of sensory stimuli. *Psychosom. Med.*, 1963, **25**, 450–458.

OBRIST, P. A., HALLMAN, S. I., & WOOD, D. M. Autonomic levels and lability, and performance time on a perceptual task and a sensory motor task. *Percept. mot. Skills*, 1964, **18**, 753–762.

RUSHMER, R. F. Autonomic balance in cardiac control. *Amer. J. Physiol.*, 1958, **192**, 631–634.

SAMAAN, A. The antagonistic cardiac nerves and heart rate. *J. Physiol.*, 1934–35, **83**, 332–340. (a)

SAMAAN, A. Muscular work in dogs submitted to different conditions of cardiac and splanchnec innervations. *J. Physiol.*, 1934–35, **85**, 313–331. (b)

WELFORD, N. T. An electronic digital recording machine: The SETAR. *J. scient. Instrum.*, 1952, **29**, 1–4.

WILSON, R. S. Autonomic changes produced by noxious and innocuous stimulation. *J. comp. physiol. Psychol.*, 1964, **58**, 290–295.

WOOD, D. M., & OBRIST, P. A. Effects of controlled and uncontrolled respiration on the conditioned heart rate response in humans. *J. exp. Psychol.*, 1964, **68**, 221–229.

ZEAMAN, D., DEANE, G., & WEGNER, N., Amplitude and latency characteristic of the conditioned heart response. *J. Psychol.*, 1954, **38**, 234–250.

ZEAMAN, D., & SMITH, R. W. Review and analysis of some recent findings in human cardiac conditioning. In W. F. Prokasy (Ed.), *Classical conditioning: A symposium*. New York: Appleton-Century-Crofts, 1964.

ZEAMAN, D., & WEGNER, N. A further test of the role of drive reduction in human cardiac conditioning. *J. Psychol.*, 1957, **43**, 125–133.

(Early publication received February 23, 1965)

Journal of Experimental Psychology
1965, Vol. 70, No. 1, 43–50

HUMAN PROBABILITY LEARNING WITH FORCED TRAINING TRIALS AND CERTAIN AND UNCERTAIN OUTCOME CHOICE TRIALS[1]

JAMES K. ARIMA

Combat Developments Command Experimentation Center, Fort Ord, California

Human Ss were used on a probability-learning task in which relative frequencies of reinforcement for 2 alternatives were .70–.30 and .90–.10 for different groups. Each of 30 choice trials (except the last) was followed by 9 forced trials where the correct alternative was identified by a distinctive cue. On choice trials, different groups at each reinforcement level were given free choices (either alternative always correct) or correction choices (one alternative correct, self-correction required). Correction-choice groups performed more or less as predicted by a probability-matching mathematical model of Estes. Free-choice groups did not "learn" to choose one alternative more frequently, but their response patterns were not random.

This study was conducted to determine what effects forced training trials interspersed among choice trials might have on two-choice probability learning. Further, the effect on choices of the correction procedure implicit in a "wrong" choice in most determinate, noncontingent, probability-learning situations was evaluated by surreptitiously letting some groups of *S*s always be correct on choice trials, while other groups of *S*s actually had to correct themselves when a wrong choice was made. Additionally, it was thought that the use of forced trials might prove to be of considerable advantage in the analysis of probability learning since it would be possible to ensure that all *S*s had made any designated number of responses to each alternative immediately preceding any choice trial, and the interspersed forced trials would separate choice trials so that stimulus traces from one choice would not directly overlap with the next choice trial.

METHOD

Let A_1 represent the response of predicting or selecting E_1, the more frequently reinforcing event; A_2 the response corresponding to E_2, the less frequently reinforcing event; and A_i and E_i these response and reinforcing events collectively.

Apparatus.—A beaver-board screen separated the apparatus visible to *S* from the apparatus visible to *E*. On *S*'s side of the screen was a plywood box with two red jewel lights and two toggle switches mounted centrally on the side facing *S*. The lights were side by side and the toggle switches were directly under the lights. The apparatus on *E*'s side of the screen was a small console which contained all the equipment necessary to select and present the stimuli to *S* and to indicate *S* s choices. A buzzer was located externally on the console and muffled. The *E* could present the buzzer alone or the buzzer simultaneously with either or both red lights on *S*'s side of the screen. The *S* could simultaneously turn off the buzzer and red light or lights by depressing only one or either of the switches on his side of the screen, as programed by *E*. A constant-speed, synchronous 10-rpm electric motor was wired to the starter circuit of E's console so that a new trial was started every 6 sec.

---

[1] This study is based on a dissertation submitted to the Graduate School, Northwestern University, in partial fulfillment of the requirements for the PhD degree. The writer most gratefully acknowledges the guidance and encouragement provided by Winfred F. Hill, chairman of his dissertation committee, and by Carl P. Duncan, Albert Erlebacher, and William A. Hunt, members.

Subjects.—The Ss were 200 students from undergraduate psychology courses at Northwestern University who volunteered for the experiment to satisfy laboratory requirements for participating in psychological experiments. They were assigned to experimental conditions, described below, in the sequence in which they appeared to be tested.

Procedure.—All Ss, following an explanation of the equipment and procedures, were given two practice trials and tested individually. Each S in the experimental groups was given 341 consecutive trials. On each trial, S's task was to guess which switch would turn off the buzzer, depress the switch he had chosen, and correct himself if the buzzer did not go off immediately. On the first 30 trials, the buzzer was presented alone without any accompanying lights, there was a 50:50 ratio of reinforcement, and the sequence was random. The same procedure was continued for 20 more trials with the exception that the red light above the correct switch was illuminated on alternate trials. The red light was illuminated on even trials for half the Ss and on odd trials for the other half. These 50 trials were given in order to bring all Ss to the same pattern of random choices before entering the 291 selective learning trials, to instill in S the expectation that only one choice would be correct on the free-choice trials that were to follow, and to permit S to learn that the switch below the illuminated light always turned off the buzzer, while the other did not. These 50 trials will hereafter be referred to as *pretraining* trials.

The 291 selective learning trials were divided into blocks of 10 trials, except for the last block which contained only 1 trial. The first trial of each block was a choice trial and the following 9 trials were forced. Since Trials 1 and 291 were choice trials, there were 30 choice trials in all. On *forced trials*, the red light was on above the switch which would turn off the buzzer. On *choice* trials, a *free-choice trial* was defined as one in which (unknown to S) the selection of either switch by S turned off the buzzer, and a *correction-choice trial* was one which required S to depress the appropriate switch which had been programed for that trial in order to turn off the buzzer. Both or neither of the red lights was illuminated on free-choice trials (see below). Neither of the red lights was illuminated on correction-choice trials.

In the comparable control conditions, all Ss were given the 50 pretraining trials followed without a break by 30 consecutive choice trials.

Groups.—There were two different probabilities (.70 and .90) for the relative frequency with which E_1 occurred and three experimental groups corresponding to each of the probabilities. One of the groups at each probability level was a free-choice group (designated FF) with two lights on choice trials, another was a free-choice group (F) without lights on choice trials, and the third was a correction-choice group (C). Thus, there were six experimental groups, constituting a 3 × 2 factorial design. In addition, there were four control groups. One control group (FFc) had both lights illuminated on its 30 consecutive free-choice trials, and another group (Fc) had no lights on its 30 consecutive free-choice trials. Two correction control groups (Cc), one at each probability level, received 30 consecutive correction-choice trials. The 10 groups and their experimental and control conditions are summarized in Table 1.

Sequence.—All sequences for the presentation of the E_i were constructed for these conditions by the use of a table of random numbers with the restriction that the exact relative frequencies of reinforcement had to be maintained within a block of 10 trials (1 choice and 9 forced) in the selective-learning stage. Three different sequences were constructed for the 50% trials, and three different sequences were constructed for each relative frequency of reinforcement. The sequences were represented approximately equally among all six experimental groups. Each S in Groups 70Cc and 90Cc had a different sequence of three consecutive 10-trial blocks selected from the three series. Except for the 50% condition, no sequences were required for the free-choice control groups, FFc and Fc.

In addition, two subsequences were constructed for each series within each reinforcement condition. These subsequences were labeled A_1 and A_2 subsequences. In the A_1 subsequence, responses on all choice trials were assumed to be A_1, and in the A_2 subsequence, it was assumed that all responses on choice trials were A_2. In the actual administration of the sequences, both sequences were used for any particular S according to the response on choice trials. For example, in the case of free-choice groups, if S's response was A_1 on the first choice trial, the sequence of forced choices from the first block of the A_1 list followed; if his second choice response was A_2 then the sequence of forced choices from the second block of the A_2 subsequence followed, etc.

TABLE 1

Experimental and Control Conditions

Group	$P(E_1)$	No. of Forced Trials	No. of Choice Trials	Type Choice Trials
70FF	.70	261	30	Free, 2 lights
90FF	.90	261	30	Free, 2 lights
70F	.70	261	30	Free, no lights
90F	.90	261	30	Free, no lights
FFc	—	0	30	Free, 2 lights
Fc	—	0	30	Free, no lights
70C	.70	261	30	Correction
90C	.90	261	30	Correction
70Cc	.70	0	30	Correction
90Cc	.90	0	30	Correction

Note.—20 *S*s per group.

The A_1 and A_2 subsequences were administered in essentially the same manner to *S*s of the correction-choice groups. The only difference was that *E* selected in advance the appropriate E_i for choice trials. Two sequences of choice trials were created for each series; one of the sequences began with E_1 on the first choice trial, and the other sequence began with E_2. The relative frequency of reinforcement prevailing on the forced trials was maintained within a block of 10 consecutive choice trials.

For half the *S*s, E_1 was on the right; E_1 was on the left for the other half.

Results

The relative frequency with which each group chose E_1 in blocks of 5 choice trials and over all 30 choice trials is shown in Table 2. Table 2 also shows the variance within groups of the number of times individual *S*s chose E_1 over the 30 choice trials. The distribution of *S*s according to the number of times they chose E_1 is shown in Table 3. The striking features of these distributions are the trimodal distributions for the free-choice groups, which is especially pronounced in the 90% groups; and the high modal values for the correction-choice groups, which are also most pronounced in the 90% groups. Chi-square tests of independence confirmed the fact that the distribution of *S*s at each reinforcement level was

TABLE 2

Relative Frequency (Percent) and Variance of Frequency of A_1 Responses on Choice Trials by Groups

Group	Blocks of 5 Choice Trials						All 30 Choice Trials	
	1	2	3	4	5	6	*M*	s^2
70FF	50	51	51	57	53	55	52.8	122.871
90FF	53	54	59	58	52	48	54.0	141.537
70F	43	37	40	49	47	50	44.3	85.168
90F	37	36	33	38	40	37	36.8	146.871
FFc	48	45	54	51	56	49	50.5	55.397
Fc	56	53	55	51	52	52	53.2	58.903
70C	55	60	57	68	69	69	63.0	20.200
90C	53	91	87	92	87	94	84.0	5.116
70Cc	59	46	58	56	58	51	54.7	15.305
90Cc	59	69	76	84	85	90	77.2	5.292

TABLE 3

Distribution of *S*s within Groups According to Number of A_1 Responses on Choice Trials

Group	No. of A_1 Responses					
	0–5	6–10	11–15	16–20	21–25	26–30
70FF	4	3	4	1	2	6
90FF	5	1	4	0	3	7
70F	5	3	3	3	4	2
90F	11	0	3	0	0	6
FFc	2	1	10	4	1	2
Fc	2	0	9	4	1	4
70C	0	1	4	6	8	1
90C	0	0	0	0	9	11
70Cc	0	0	8	9	2	1
90Cc	0	0	0	2	16	2

dependent on the type of choice procedure used. Accordingly, further analyses of results were conducted separately for the correction and free-choice groups.

Correction-choice groups.—Figure 1 presents the performance curves of the correction-choice groups (70C and 90C) and their respective control groups (70Cc and 90Cc). The smooth

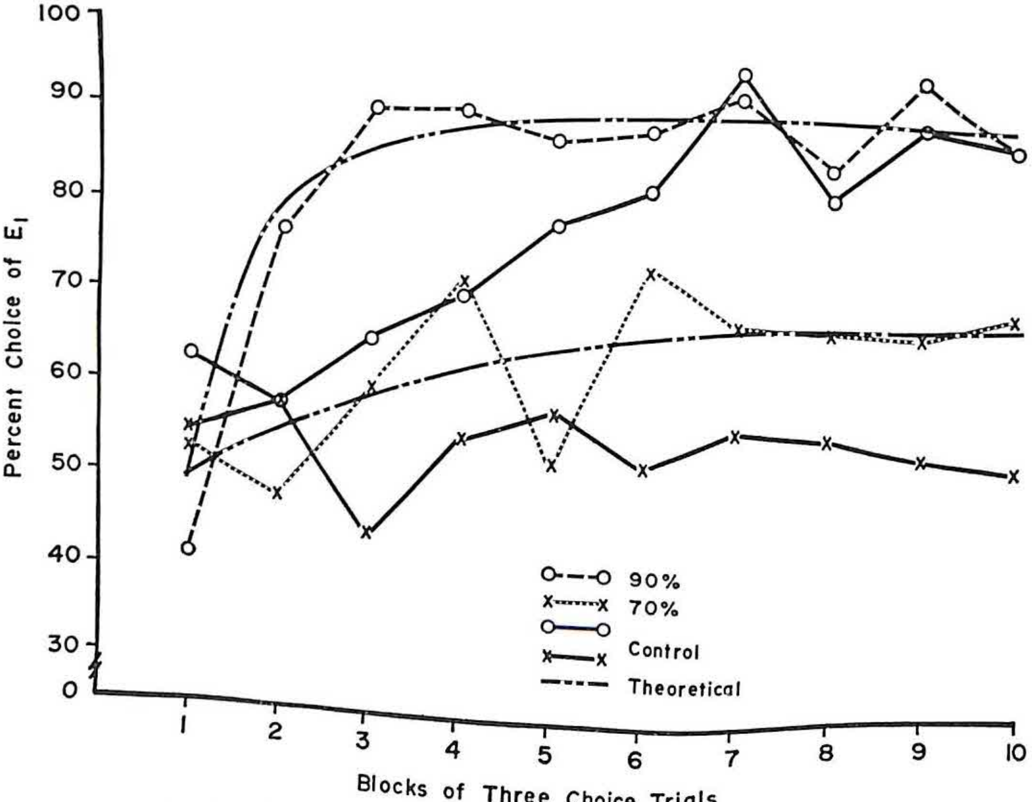


Fig. 1. Obtained and theoretical curves of the percent choice of the E_1 event by correction-choice groups.

curves were generated by an equation given by Estes (1959, p. 411). The value of P_n at the beginning of selective learning was taken to be 50% in both cases, since the effect of the pretraining trials was to bring the mean probability of an A_1 response to .542 for *Ss* in both experimental groups combined. The value of θ for Group 70C was found to be .095 and that for Group 90C was .222. The equation for calculating the smooth curve of Group 70C was

$$P_m = .7 - (.2)(.905)^{3m-1}$$

and that for Group 90C was

$$P_m = .9 - (.4)(.778)^{3m-1},$$

where m is the ordinal number of a block of three trials.

An analysis of variance for trends (Edwards, 1960) of the correction groups is presented in Table 4. The results were broken down into six blocks of five choice trials for this analysis. The main effects of reinforcement and training are significant in both their linear and quadratic components. The significant interaction of reinforcement with blocks results because the learning curve for the combined 90% groups is more accelerated than the curve for the combined 70% groups, and because the configuration of the two curves differs. The significant interaction of training with blocks arises almost solely from differences in curve configurations between the combined experimental and combined control groups and from the fact that both of these learning curves show slight reversals from a monotonic growth function. The triple interaction is not significant.

TABLE 4

ANALYSIS OF VARIANCE (TRENDS) OF FREQUENCY OF A_1 RESPONSES FOR CORRECTION GROUPS

Source	*df*	*MS*	*F*
Between *Ss*	(79)		
Reinforcement (R)	1	141.919	74.851***
Training (T)	1	17.252	9.099**
R × T	1	.169	—
Ss/Groups	76	1.896	—
Within *Ss*	(400)		
Blocks (B)	5	11.112	11.784***
Linear	1	46.629	50.464***
Quadratic	1	7.668	8.389**
Residual	3	.421	—
R × B	5	6.739	7.146***
Linear	1	16.830	18.214***
Quadratic	1	5.316	5.816*
Cubic	1	9.870	13.410***
Residual	2	.838	—
T × B	5	2.922	3.099**
Linear	1	1.821	—
Quadratic	1	1.768	—
Cubic	1	5.483	7.450**
Quartic	1	2.511	—
Residual	1	3.027	—
R × T × B	5	1.229	—
Ss/Groups × B	380	.943	
Linear	76	.924	
Quadratic	76	.914	
Cubic	76	.736	
Quartic	76	1.032	
Residual 1	228	.960[a]	
Residual 2	152	1.072[a]	
Residual 3	76	1.112[a]	
Total	479		

[a] Residual 1 is the *Ss*/Groups × B sum of squares less those for linear and quadratic components; Residual 2 is the *Ss*/Groups × B sum of squares less those for linear, quadratic, and cubic components; Residual 3 is the *Ss*/Groups × B sum of squares less all component sums of squares.

* $p \leq .05$.

** $p \leq .01$.

*** $p \leq .001$.

The effect of forced training trials on *S*'s response on any Choice Trial n was evaluated by computing the conditional probabilities between Forced Trial $n-1$ and *S*'s choice on Trial n. Sequential dependencies in *S*'s choices and the effect of a choice and its outcome on a succeeding choice were evaluated by computing the conditional probabilities between the joint events, A_iE_i, on Trial $n-10$ and the choice on Trial n. Only the effects of a preceding forced or choice trial were evaluated since previous research suggests that the information in *S*'s response is related, primarily, to the immediately preceding event (Anderson, 1960; Jarvik, 1951).

Conditional probabilities will be written in the form, $P(A_i, E_i)$, and will be read, "The probability of an A_i, given an E_i." Joint events will be written in the form, A_iE_i, and the probability of a joint event will be written, $P(A_iE_i)$. The conditional probability between a joint event and a succeeding choice will be written, $P(A_i, A_iE_i)$, and will be read, "the probability of A_i, given the joint event A_iE_i."

The conditional probabilities between a given reinforcing event on Trial $n-1$ and a choice on Trial n are shown separately in Fig. 2 for the 70 and 90% groups according to successive thirds of the selective learning trials. The $P(A_1, E_1)$ are shown by solid lines and the $P(A_1, E_2)$ are shown by dashed lines. The corresponding curves of $P(A_2, E_1)$ and $P(A_2, E_2)$ are complements of these curves and are not shown.

The conditional probabilities between a given joint event on Trial $n-10$ and an A_1 response on Trial n are shown in Fig. 3 for Group 70C and in Fig. 4 for Group 90C. The conditional probabilities of A_2, given these joint events, are not shown since they, too, are complements of the curves shown.

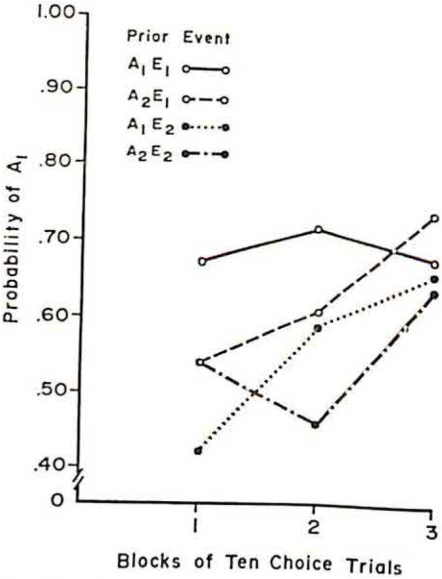


FIG. 3. Curves of the conditional probability of A_1 following the joint occurrence of response and reinforcement (A_iE_i) on choice trials in Group 70C.

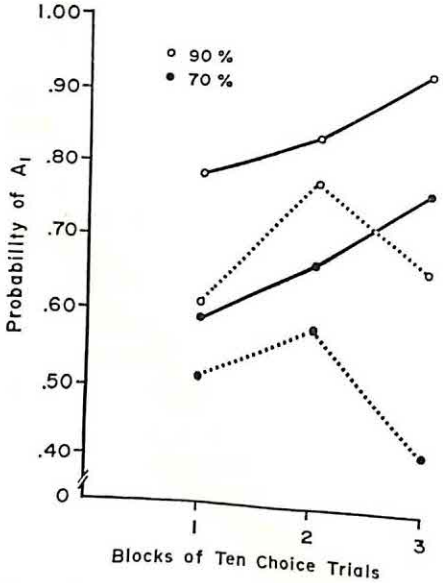


FIG. 2. Curves of the conditional probability of A_1 following forced trials on which E_1 (solid lines) and E_2 (dashed lines) occur in correction-choice experimental groups.

Free-choice groups.—It was felt justified on the basis of evidence presented by Boneau (1960) to conduct a parametric analysis of the data

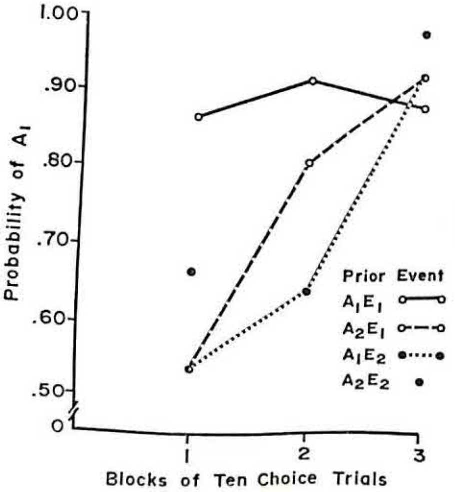


FIG. 4. Curves of the conditional probability of A_1 following the joint occurrence of response and reinforcement (A_iE_i) on choice trials in Group 90C. (No joint events A_2E_2 occurred in the second block of trials.)

for the free-choice groups. An analysis of variance for trends, similar to that for the correction groups, was made for the six free-choice groups. The main variables were reinforcement and lights (the presence or absence of red lights on choice trials). The effect of training (interspersed forced trials) vs. no training was not considered as a separate dimension since there seemed to be no consistent relationship on this variable. The control groups were therefore included in the analysis as a level of reinforcement. The effects of lights or no lights on free-choice trials, on the other hand, appeared as if it might possibly be significant. Thus, there were three levels of reinforcement—"control," 70, and 90%—for the analysis, and at each level of reinforcement there was one group which had two lights on free-choice trials and another which had no lights on its free-choice trials. Neither the main effects nor their interactions were statistically significant. The effect for blocks was very small ($F = 1.133$) and none of the interactions of the main effects with blocks reached statistical significance. Separate *t* tests were conducted for each of the experimental groups against the null hypothesis that the mean number of A_1 responses per *S* over all 30 choice trials did not differ from a random (50%) choice of the E_1 alternative. In no case could the null hypothesis be rejected at the .05 level of probability.

The conditional probabilities between a response class on a choice trial and either reinforcing event on the preceding forced trial was merely the probability of the response.

Since any choice on a choice trial was reinforced with the corresponding event for the free-choice groups, the A_i and the E_i were always the same on any choice trial, and the joint events that could occur on any choice trial were A_1E_1 and A_2E_2. Therefore, the conditional probability between the occurrence of a joint event on Choice Trial $n-10$ and *S*'s choice on a succeeding Choice Trial *n*, can be stated as the probability of a repetition. These were calculated for all experimental groups by blocks of 10 trials, and it was found that the probabilities for the repetition of A_1 and A_2 responses were independent of each other and reflected in the same way a general trend toward more or less stereotopy in responding. These probabilities were within a range of .60–.80 for the 70% groups and within a range of .70–.90 for the 90% groups (with one minor exception).

DISCUSSION

It would appear that the statistical-learning model used here adequately predicts the general shape of the overall mean learning curves and the asymptotic level of responding of the correction-choice groups even when most of the training trials are forced and learning can be considered a transfer task from an acquisition task with 50% reinforcement of each alternative. Systematic trends between preceding events and responses on choice trials are observed, however, in Fig. 2, 3, and 4. Early in training, *S*'s choice depends considerably on the choice and outcome on an earlier choice trial as revealed by the differing probabilities of A_1 in Fig. 3 and 4, given the various joint events on Trial $n-10$. It is only in the last third of training in Fig. 3 and 4 that the probabilities of A_1 become relatively independent of the joint events on Trial $n-10$. However, in the last third of training, *S*s also begin to rely more on the immediately preceding event programed by *E* in making a new choice. The curves in Fig. 2 suggest that A_1 responses begin to occur with a probability greater than $P(A_1)$ when the preceding event is E_1, and the probability of A_1 responses actually begins to decline when the preceding event is E_2. Eventually, it may be that *S*s go through other changes in strategy until they give up trying to predict the less frequent event and maximize hits by giving only A_1 responses (Edwards, 1961).

It was shown that the presence or absence of lights on choice trials made no difference for the free-choice groups. In either case, they seemed to show no "learning." Therefore, the succeeding discussion will ignore the differences in lights on choice trials for the free-choice groups. Since both the correction and free-choice groups were treated identically with respect to pretraining and forced trials, their differences in performance can only be attributed to differences in treatment on choice trials. The effect of different choice procedures, devoid of the influence of forced trials, can be seen in the performance of the control groups. Except for rate, correction-choice control groups showed learning trends similar to the experimental correction-choice groups. The free-choice control groups, like the experimental groups, showed no learning.

Closer inspection of the performance of the 40 free-choice control *S*s revealed that 10 *S*s perseverated position responses and 10 *S*s perseverated patterns of single, double, and triple alternations over the 30 trials. Such persistent strategies were not seen in the correction-choice control groups.

In contrast with the performance of the free-choice controls, only one of the 80 *S*s in the four free-choice experimental groups adopted and maintained an alternation pattern on choice trials. On the other hand, there was a great tendency for *S*s in the free-choice experimental groups to repeat choices (Table 3). Thus, the effect of the forced trials was to eliminate alternation tendencies and strengthen the tendency to repeat responses. This trend was more evident in the 90% groups.

In conclusion, it can be stated that a probability-matching type of learning took place when correction procedures were used on choice trials and that this learning was more rapid when forced training trials were interspersed among the choice trials. When *S* was always correct on choice trials, no preferential choice of any alternative was shown by groups, even when forced training trials were interspersed among the choice trials.

REFERENCES

Anderson, N. H. Effect of first-order conditional probability in a two-choice learning situation. *J. exp. Psychol.*, 1960, **59**, 73–93.

Boneau, C. A. The effects of violations of assumptions underlying the *t* test. *Psychol. Bull.*, 1960, **57**, 49–64.

Edwards, A. L. *Experimental design in psychological research.* New York: Holt, Rinehart, & Winston, 1960.

Edwards, W. Probability learning in 1000 trials. *J. exp. Psychol.*, 1961, **62**, 385–394.

Estes, W. K. The statistical approach to learning theory. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. New York: McGraw-Hill, 1959. Pp. 383–491.

Jarvik, M. E. Probability learning and a negative recency effect in the serial anticipation of alternative symbols. *J. exp. Psychol.*, 1951, **41**, 291–297.

(Received March 24, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 1, 51–56

PAIRED-ASSOCIATES LEARNING WITH VARYING RELATIVE PERCENTAGES OF OCCURRENCE OF ALTERNATIVE RESPONSE MEMBERS: INFLUENCE OF INSTRUCTIONS[1]

ALBERT E. GOSS

University of Massachusetts

Acquisition and performance with divergent PA units were determined for 4 relative percentages of occurrence of alternative response members (RP) combined with 4 list conditions under matching, regular, and maximizing instructions. Under all 3 instructions and irrespective of RP, percentages of anticipations and actual occurrences of more frequent response members were essentially the same for a single unit presented at constant, short, and at variable, long interunit intervals. With RP of 70-30, 80-20, and 90-10, for a single unit presented with randomly paired intervening stimuli and for 4 units, matching instructions reduced differences between anticipations and actual occurrences. Maximizing instructions increased those differences but did not occasion 100% anticipation.

Considered together, the three experiments reported here provide information about effects of different sets of instructions on acquisition and performance with divergent paired-associates (PA) units. These units were each constituted of a stimulus member and two response members, one of which accompanied the stimulus member on $P\%$ of its occurrences, the other of which accompanied the stimulus member on $(1 - P)\%$ of its occurrences. Relative percentages of occurrence of alternative response members (RP) were varied as were list conditions which represented differences in number of units and manner of presentation of single units.

One reason for varying instructions was empirical: to determine direction and extent of effects of different sets of instructions on acquisition and performance both across RP and list conditions and for different RP and list conditions alone and in combination. Another reason for varying instructions was theoretical: to determine whether or not theories to explain acquisition and performance with divergent PA units need be elaborated to allow for effects of instructional variables. The n-choice or probability-learning task (Estes, 1964) can be considered a special case of divergent PA units. Therefore, findings of effects of instructional variables on acquisition and performance with divergent PA units might have interesting and important ramifications for theories of probability learning.

METHOD

Design.—Three different sets of instructions were combined orthogonally with RP of 60-40, 70-30, 80-20, and 90-10 and with list conditions of 4 units or 1 unit and three different manners of presentation of the single unit. The three different manners were with a constant 2-sec. interval between each presentation of the single unit; with intervals from 2 to 38 sec. between presentations about a mean of 20 sec.; and with intervals from 2 to 38 sec. between presentations about a mean of 20 sec., but with the intervals filled by presentations of randomly paired stimulus and response members. These different manners of presentation are described in greater detail by Goss and Sugerman (1961). Coded by them as 1 unit, 1 unit spaced, and 1 unit random, they are, henceforth, coded here as 1, 1S, and 1R, respectively. The 4-unit list is coded here as 4.

---

[1] Supported by Contract 2691(00) between the Office of Naval Research, Training and Personnel Branch, and the University of Massachusetts. Marilyn E. Sugerman collaborated on one experiment. Nancy C. Farrick was responsible for running half of the *S*s of that experiment and all of those of the other two. She and Helen L. Rowell contributed to analyses of the data of the latter experiments and of all three experiments combined as well as to preparation of the manuscript.

The three different sets of instructions are referred to as regular, matching, and maximizing. The regular instructions were those employed for acquisition with conventional PA units of each stimulus member accompanied by but one response member. The *S*s were not informed that each stimulus member would be accompanied by two alternating response members. Nor were they told the manner in which they should anticipate those members except to spell out the stimulus they expected to occur before the response member was exposed.

The matching instructions involved informing *S*s that each stimulus member would be accompanied by two response members which would alternate. The *S*s were told further to anticipate the response members with frequencies which were the same as the actual frequencies of occurrence of those members. The maximizing instructions also involved informing *S*s that each stimulus member would be accompanied by response members which would alternate. They were told further, however, to always anticipate the more frequent response member of a unit.

The design was accomplished in three successive experiments. The first experiment, with regular instructions, has been described previously (Goss & Sugerman, 1961). Once acquisition and performance with divergent PA units under regular instructions had been determined for RP and list conditions alone and in combination, comparisons with acquisition and performance under other sets of instructions were warranted and desirable. The second experiment was with matching instructions and the third experiment was with maximizing instructions.

Goss and Sugerman varied RP from 50-50 to 100-0 in successive changes of 10%. Since instructions to match or to maximize are of little or no consequence for RP of 50-50 or 100-0, these levels were omitted in the two later experiments. Except for the omission of RP of 50-50 and 90-10, the only known differences between the first experiment and the two later experiments were in the times the experiments were run and in *E*s. The former experiment was run during the spring of one year, the experiment with matching instructions was run in the fall of that year, and the experiment with maximizing instructions was run in the spring of the next year. There is no reason to believe that *S*s of the three experiments, who were drawn from among undergraduates enrolled in the course in introductory psychology during those semesters, differed in any ways related to acquisition with divergent PA units.

The *S*s in the Goss and Sugerman experiment were run by two female *E*s. No differences attributable to *E*s were observed. In the two later experiments, one of these *E*s ran all of the *S*s.

Lists.—The divergent PA units were constituted of CVCs of high meaningfulness and low formal similarity to each other. The units were WOM accompanied by TEX or LIK, RUF by CEN or DOL, FES by VIC or BAL, and NAV by HOB or SUR. When presented as four units, TEX, DOL, VIC, and HOB all occurred on 60%, 70%, 80%, or 90% and LIK, CEN, BAL, and SUR all occurred on 40%, 30%, 20%, or 10% of the presentations of their stimulus members, or the converse. When presented as 1 unit, which response member was the more frequent was also counterbalanced. Goss and Sugerman (1961) describe other details of the lists and of the semirandom orders of presentation of the units.

Apparatus and procedure.—A Hunter cardmaster was employed to present the units; the rate was 2:2 sec. There were 120 trials with the 4-unit lists with all units presented on each trial; there were 120 presentations of the single units under the 1, 1S, or 1R list conditions.

Subjects.—Within each experiment there were 16 combinations of RP and list conditions. These combinations were administered in counterbalanced cycles until the requisite number of 8 *S*s within each had been reached. Assignment to each combination was in order of *S*s' appearance within a constraint of equal numbers of males and females within each combination. There were 128 *S*s within each experiment, for a total of 384. All *S*s were naive with respect to exposure to lists of divergent PA units. Some *S*s within each experiment, however, had served previously in other experiments with conventional PA units.

RESULTS AND DISCUSSION

As previously, the response measure was percentage of anticipations of the

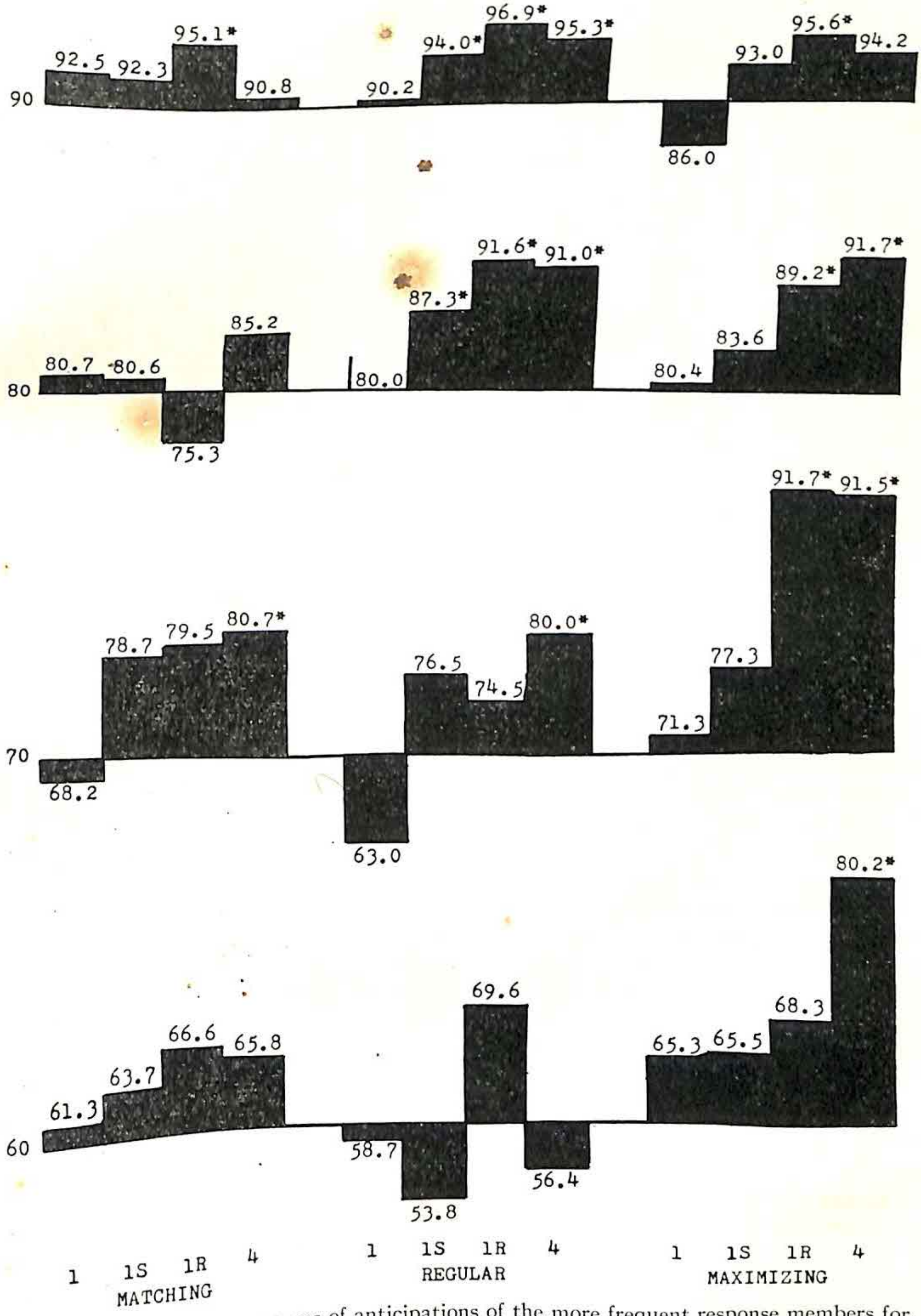


FIG. 1. Means of percentages of anticipations of the more frequent response members for each of the combinations of instructions and lists in relation to actual percentages of occurrence of the more frequent response members. (Differences significant at $p \leq .05$ are indicated by an asterisk.)

TABLE 1

ANALYSES OF VARIANCE ON PERCENTAGES OF ANTICIPATIONS OF THE MORE FREQUENT RESPONSE MEMBERS DURING ALL SIX AND THE LAST THREE BLOCKS OF TRIALS

Source	All Six Blocks			Last Three Blocks		
	df	*MS*	*F*	*df*	*MS*	*F*
Instructions (I)	2	2,458.08	3.97*	2	2,232.75	7.11**
RP	3	89,733.75	145.09**	3	41,646.13	132.61**
List conditions (L)	3	2,313.34	3.74*	3	4,690.09	14.93**
I × RP	6	1,929.71	3.12**	6	1,302.86	4.15**
I × L	6	1,049.72	2.15	6	581.73	1.85
RP × L	9	767.16	1.57	9	553.78	1.76
I × RP × L	18	488.34		18	334.82	1.07
Error (b)	336	618.47		336	314.04	
Trials (T)	5	37,680.46	327.20**	2	360.30	6.17**
T × I	10	197.79	1.72	4	45.41	
T × RP	15	161.09	1.40	6	90.05	1.54
T × L	15	3,039.13	26.39**	6	119.23	2.04
T × I × RP	30	86.36		12	42.86	
T × I × L	30	158.70	1.38	12	56.32	
T × RP × L	45	172.26	1.50*	18	30.04	
T × I × RP × L	90	117.05	1.02	36	85.91	1.47*
Error (w)	1672[a]	115.16		664[a]	58.38	

[a] 8 estimated values subtracted from *df*.
* $p \leq .05$.
** $p \leq .01$.

more frequent response members during successive blocks of 20 trials. For brevity, means of percentages for successive separate blocks of trials are not shown. Means of percentages during the last three blocks of trials combined are presented at the top of each rectangle of Fig. 1. Means of percentages of anticipations are first examined in terms of differences among their absolute values through blocks of trials as functions of instructions, RP, and list conditions. Direction and extent of differences between anticipations of response members and their actual occurrence are then examined.

Absolute values.—Table 1 summarizes analyses of variance on percentages of anticipations during each of all six and the last three blocks of trials. In each analysis, the *F* for instructions was significant. Both across all six and the last three blocks of trials, differences between anticipations under matching and regular instructions were slight. Fewer anticipations occurred under both of these sets of instructions than under maximizing instructions. Both of the highly significant *F*s for RP reflected the direct relationship between percentages of anticipations and of actual occurrences of the more frequent response members found in other experiments (Brown, Osterhout, & Voss, 1962; Voss, Thompson, & Keegan, 1959). The significant *F*s for list conditions are considered below in connection with effects of trials.

The *F*s for I × RP were also significant. For all six and for the last three blocks of trials, percentages of anticipations were in the order regular < matching < maximizing with RP of 60-40 and 70-30, in the order matching < maximizing < regular with RP of 80-20, and in the order maximizing < matching < regular with RP of 90-10. None of these orders is the same as the overall order for instructions of matching ≐ regular < maximizing.

Neither of the *F*s for I × L was significant. Across all six blocks of trials, maximizing instructions produced more anticipations than did regular or matching instructions for all but the 1 list condition. Across the last three blocks of trials, maximizing instructions produced more anticipations than did matching and regular instructions for all four of the list conditions. Thus, while the higher percentages of anticipations under maximizing than under matching and regular instructions suggested by the overall *F*s for instructions did not hold at each RP, with the one exception, the higher percentages did hold for each list condition.

Examination of the rank orders of means for each of the 16 combinations of RP and list conditions under the different instructions also indicated that maximizing instructions did, in general, produce more anticipations than did regular or matching instructions. However, neither of the *F*s for I × RP × L was significant.

The highly significant *F* for trials through all six blocks reflected a typical negatively accelerated increase in percentage of anticipations of the more frequent response members. The *F* for trials through the last three blocks was also significant. More anticipations occurred during Trials 101–120 than during Trials 61–80 for which the mean was slightly higher than the mean for Trials 81–100. However, none of these differences was sufficient to be of practical importance or to vitiate the subsequent treatment of the last 60 trials as essentially homogeneous.

The significant T × L interaction for all six blocks was attributable to higher percentages of anticipations with the 1 and 1S list conditions than with the 1R and 4 list conditions during Trials 1–20, and the converse during the last four blocks of trials. The pattern of this interaction substantially accounts for the difference between the orders of the list conditions across all six and across the last three blocks of trials. Across all six blocks, the initial advantage of the 1S list condition was sufficient to produce the order 1 < 1R < 4 < 1S. During the last three blocks, however, the order was 1 < 1S < 1R ≐ 4. Both *F*s for list conditions were significant.

The significant T × RP × L interaction reflected a greater reversal in the advantage of the 1 and 1S to the 1R and 4 list conditions between Trials 1–20 and the last four blocks of trials with RP of 60-40 and 70-30 than with RP of 80-20 and 90-10.

For the last three blocks of trials, only one interaction involving trials was significant. However, the pattern of changes through trials precluded meaningful interpretation of this interaction.

Anticipations and actual occurrences.—Figure 1 indicates the direction and extent of differences between percentages of anticipations of response members and their actual occurrences. The heterogeneity of the variances among the combinations of instructions, RP, and list conditions was not considered an important limitation on interpretation of the outcomes of the analyses of variance. But such heterogeneity seemed an important limitation on assessment of differences between percentages of anticipations and of actual occurrences. Accordingly, a two-tailed *t* (7) was computed for each of the 48 combinations with an error term based on the *SD* for that particular combination. Differences significant at $p \leq .05$ are noted in Fig. 1.

Under all three sets of instructions, only 2 of the 24 *t*'s for RP combined with 1 and 1S list conditions were significant. Under matching instructions, 2 of the 8 *t*'s for RP combined with 1R and 4 list conditions were significant. In contrast, under regular and maximizing instructions, respectively, 5 and 6 of the 8 *t*'s for RP combined with 1R and 4 were significant.

The pattern of nonsignificant and significant *t*'s, in conjunction with the outcome of the analysis of variance for the last three blocks of trials, supports four conclusions. First, regardless of instructions, during the last three blocks of trials percentages of anticipations for combinations of RP with 1 and 1S list conditions matched actual percentages of occurrence. Two significant *t*'s among 24 are best interpreted as due to chance. Second, regardless of instructions, percentages of anticipations for RP of 60-40 combined with 1R and 4 list conditions also matched actual percentages of occurrence. Because of the extent of the increase in the percentage of anticipations over 60%, the combination of RP of 60-40 and the 4 list condition under maximizing instructions may be an exception.

Third, instructions to match may have reduced differences between anticipations and actual occurrences for RP of 70-30, 80-20, and 90-10 combined with the 1R and 4 list conditions. Instructions to maximize may have increased those differences for the same combinations. The largest absolute differences were under instructions to maximize for RP of 60-40 combined with the 4 and of 70-30 combined with the 1R and 4 list conditions. Thus, theories of learning and performance with divergent PA units, including theories of probability learning, may have to be elaborated to allow for effects of different sets of instructions.

The Group 2 and Group 6 instructions employed by McCracken, Osterhout, and Voss (1962) resembled the matching and maximizing instructions of the present experiments, respectively. Their task approximated the 1 list condition of the present experiments; the RP was 70-30. Under Group 2 instructions, percentages of anticipations of the more frequent stimulus were close to 70%. Under Group 6 instructions, percentages of anticipations rose to about 90%. In contrast, with 1 and 1S list conditions under maximizing instructions, anticipations matched actual occurrences at each level of RP. One or more of many differences between the McCracken et al. and the present experiments in lists and procedures might underlie the discrepant outcomes. Thus, for example, maintenance of instructions-induced sets may have been less difficult with their simple nonmechanical arrangement for presenting the stimuli than with the complicated, noisy cardmaster of the present experiments.

Fourth, regardless of any differences in effects of regular and maximizing instructions, under these instructions anticipations tend to exceed actual occurrences for RP for 70-30, 80-20, and 90-10 combined with the 1R and 4 list conditions.

REFERENCES

Brown, A., Osterhout, C., & Voss, J. F. The effect of one probabilistic verbal association upon the concomitant acquisition of a second association. *J. verbal Learn. verbal Behav.*, 1962, **1**, 48–53.

Estes, W. K. Probability learning. In A. W. Melton (Ed.), *Categories of human learning.* New York: Academic Press, 1964.

Goss, A. E., & Sugerman, M. Paired-associates learning with varying relative percentages of occurrence of alternative response members. *J. exp. Psychol.*, 1961, **62**, 24–34.

McCracken, J., Osterhout, C., & Voss, J. F. Effects of instructions in probability learning. *J. exp. Psychol.*, 1962, **64**, 267–271.

Voss, J. F., Thompson, C. P., & Keegan, J. H. Acquisition of probabilistic paired associates as a function of S-R_1, S-R_2 probability. *J. exp. Psychol.*, 1959, **58**, 390–399.

(Received April 6, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 1, 57-62

COHERENCE DETECTION: A BASIC MECHANISM[1]

ALBERT S. RODWAN

Emory University

This study was designed to test 2 hypotheses about perception: the subjective-standards hypothesis and a coherence-detection theory. The stimulus set consisted of 100 plane figures. 3 Ss were required to decide whether the stimulus was a square or a rectangle, or later whether it was a tall rectangle or a short rectangle. 10 other Ss were required to decide whether the stimulus was a tall rectangle, a square, or a short rectangle. The results showed that (a) Ss were reliable, (b) the basis for the judgments was identical, (c) a vertical-horizontal illusion exists, and (d) there was no significant difference between the mean subjective square and the mean subjective short rectangle. The inference drawn is that "square" is a response label and not a perceptual category.

Hake, Rodwan, and Weintraub (in press), and Rodwan and Hake (1964), and Rodwan (1964) have offered evidence for the argument that perception is a coherence-detecting system. Coherence detection is a technique used by communication receivers which enables them to separate signals from noise whenever a correlation can be established between some components of a noisy signal and a standard of comparison. In perception, one such standard may be a subjective one, i.e., a memory, and the correlation is between the presented stimulus set and a set of perceptual (memory) categories. On the other hand, the correlation among different components of the stimulus objects may produce a standard. The second position denies the need for memory for perceptual categories, at least insofar as perception is concerned.

William James (1963) in his abridgement of the two-volumed *Principles of Psychology*, saw no need to modify "a general law" of perception, which has its adherents today. He said "*while parts of what we perceive come through our senses from the object before us, another part* (and it may be a larger part) *always comes out of our mind* [italics his] [p. 294]." He pointed out that there were no "pure sensations" and that perception was the interaction of the presented object and memory, association, and the like. Eriksen and Hake (1957) called this position the "subjective-standard hypothesis" and said that perception never occurs in a vacuum, that it is always relative, and that even though a stimulus may be presented individually, in fact, *O* makes a comparative judgment: he compares the presented stimulus with a subjective standard.

Two points of view can thus be contrasted: absolute judgments in perception are comparative in that the presented stimulus is compared with a subjective standard (a memory, schema, template, etc.); or, the underlying mechanism for absolute judgments in perception is coherence detection, a process which is sensitive only to relationships among the attributes of the stimulus object.

The following experiment was designed to demonstrate the existence of a situation which cannot be handled by the subjective-standard hypothesis.

---

[1] This study was supported in part by Grant GB1249 from the National Science Foundation.

METHOD

Subjects.—The *S*s were three undergraduate psychology majors, one male and two females. A requirement was that each have at least 20/20 vision in the right eye. This testing was done by means of a Bausch and Lomb orthorater.

Apparatus.—The stimuli for this experiment were presented by means of a specially constructed three-arm tachistoscope made by the Ridgely Company, Champaign, Illinois. The maximum error for a square rectangle generator, located in the middle arm is .03 mm. The stimulus figure, a rectangle in a dark surround, was backlighted by two 4-w. fluorescent bulbs and was presented at a distance of 36 in. The fixation point is centered within this area. Only the right eye was used.

Stimuli.—The set of stimuli were: (*a*) One hundred plane figures, the dimensions of which varied in height and width by .1 mm. (from 50.8 mm. to 51.70 mm.). The 10 dial settings for height and the 10 for width were combined so that 100 stimuli were presented. (*b*) One hundred plane figures, the dimensions of which varied by .2 mm. (from 50.8 mm. to 52.7 mm.). As in *a*, there were 10 dial settings for height and 10 for width.

Three combinations of judgment categories were used: A—tall rectangle, square, short rectangle; B—rectangle, square; C—tall rectangle, short rectangle. Each combination of categories was paired with *a* and *b*, thus generating six conditions. Each of the six conditions was repeated twice. Due to a misunderstanding, however, and an error on the part of *E*, the A conditions were not recorded. A second experiment was performed on the A condition. The proposed analysis was performed on the two B conditions as well as on the two C conditions. There were two orders of presentation of the conditions. The two female *S*s viewed the stimuli with the conditions ordered: $C_1\ B_1\ A_1\ A_2\ B_2\ C_2$; $A_3\ B_3\ C_3\ C_4\ B_4\ A_4$, and the male when the conditions were ordered: $A_1\ B_1\ C_1\ C_2\ B_2\ A_2$; $C_3\ B_3\ A_3\ A_4\ B_4\ C_4$.

Procedure.—The *S*s were told that they were participating in an experiment in perception involving squares and rectangles and that they would be asked to make judgments of the stimuli presented to them by means of the tachistoscope.

A single condition was run each day for 12 days. At the beginning of each condition, *S* was instructed as to which categories he was to use in making the judgments of the stimuli, i.e., Cond. B or C.

One hundred and fifty stimuli were presented in each session, the first 50 being practice trials. As *S* viewed the fixation point, *E* set the dials for the particular stimulus to be shown. When the dials were set, *E* called aloud the number of the stimulus being shown and at the same time depressed the switch for the presentation of the stimulus. The stimulus remained on for 1 sec. At the end of the presentation, *S* called aloud his judgment and recorded it on a data sheet, indicating the category chosen with the letter, SR, TR, or SQ, depending upon the conditions of the day's trials. The *S* then returned his attention to the fixation point as *E* set the dials for the next stimulus. The average intertrial time was 9 sec.

Analysis.—The first type of data analysis in this study was in terms of *d′* (Swets, Tanner, & Birdsall, 1961). This is a measure of recognition which is relatively independent of response bias and response rates. The *d′* measure assumes that *S* tries to maximize his ability to correctly recognize stimuli and at the same time take into account the values and costs of his different types of errors and hits (Swets et al., 1961).

The second type of analysis was with the linear discriminant function (LDF). The LDF is the behavioral analogue of a subjective decision axis and assumes that (*a*) *S* has two response categories to which he can assign the stimuli; (*b*) a stimulus is a point in a *p* space if it has *p* psychological attributes; and (*c*) *S* assigns stimuli to categories in such a way that a weighted linear function best describes his decision axis, i.e., his judgmental process. This weighted linear function is the LDF and it is that straight line which maximizes discriminability between the two categories (see, Rodwan & Hake, 1964).

RESULTS

Experiment I.—The *d′* measures were computed for each replication, each judgment and each *S*, and then averaged. The mean *d′* for the .1-mm. increment was .07 and for the .2-mm. increment was .01. The *d′* was computed from a matrix whose rows were *S*s' uses of the categories "square" and "rectangle" and whose columns were "square" and "rectangle" defined physically. This, then, is a measure of the agreement between *S*'s use of the responses and the physically de-

TABLE 1

Test-Retest Reliability Coefficients within the Two Conditions Uncorrected for Attenuation

	B_1	B_2	C_1	C_2
PH	.97	.87	.98	.94
LP	.96	.96	.99	.99
TS	.90	.99	.97	.99

fined categories. Thus *S*s could not recognize squares and rectangles when defined physically. Nevertheless, the results show that each *S* was responding reliably to the stimulus set (this has been demonstrated previously; see Rodwan, 1964.)

Table 1 presents the reliability coefficients separated by *S* and condition. For each *S* and each session a discriminant function was derived and was used to give a numerical value to each stimulus figure. Table 1 presents the correlation between the discriminant values for the same judgments at two different times. These coefficients, then, become test-retest reliability coefficients. The lowest reliability (.87) is for PH in the square-rectangle condition, when the increment between the adjacent stimuli was .2 mm., i.e., twice as large as the increment in the first column. All the other coefficients are large enough to be accepted as unity. Table 1 indicates that, in general, increasing the "distance" separating stimuli has no effect on reliability even though a .1-mm. increment is too small to produce a "correct" judgment square rectangle, *O* is quite reliably doing something. As a result, the data for the .1-mm. and .2-mm. increments were pooled for each *S*.

The proportion of the variance, r^2, in the square rectangle condition (B) which can be accounted for on the basis of judgments in the tall rectangle, short rectangle condition (C) is .998 for PH, .978 for CP, and .999 for TS. That is to say, a discriminant function was computed for the square rectangle judgments and for the tall rectangle, short rectangle judgments. Then the discriminant function for the tall rectangle-short rectangle was used to predict which category *S* would assign the stimulus figures to in the square rectangle judgments. This is a "cross-prediction." In the 200 judgments obtained under Cond. B (square rectangle), Cond. C (tall rectangle-short rectangle) discriminant function was incorrect in predicting at most three judgments for LP. This is very strong evidence that *S*s were "doing the same thing" in both judgments, even though, on a priori grounds, it would appear that they were different judgments.

Table 2 gives the set of computed T^2 values in testing for the existence of a vertical-horizontal illusion. This test was first presented by H. Hotelling and has been offered many times since (Anderson, 1958). As used here the test was for the differences between the mean subjective square, as given by *S*, and the geometric square. A positive vertical-horizontal illusion would occur when the mean for the subjective square was shorter than "it ought to be." All the T^2 values in Table 2 are significant at the .01 level. It is interesting to note that T^2 values computed for the condition where an increment of .1 mm. was used were

TABLE 2

T^2 Values Computed in the Square Rectangle Condition

	PH	LP	TS
B_1	12.46	17.05	30.97
B_2	12.01	7.89	6.35

Note.—$df = 2/97$. A value of 4.88 is significant at the .01 level.

consistently larger than those for the .2-mm. increment. A check was made of this by running five more *S*s only under this condition. In these five, there were three *S*s who produced a larger T^2 under a .2-mm. increment. Care, therefore, must be exercised in that it is not appropriate to decide from Table 2 that the vertical-horizontal illusion was greater for a .1-mm. increment. It does appear reasonable to infer that when an increment separating adjacent stimuli is either .1 mm. or .2 mm. (or between) there will be a measurable illusion. The values of the T^2 test in a comparison between the mean subjective "square" in Cond. B with the mean subjective "short rectangle" in Cond. C were: T^2 (2, 77) = 2.01 for PH, T^2 (2, 80) = 3.05 for LP, and T^2 (2, 61) = 2.61 for TS. All of these T^2 values were not significant. This test allowed only one decision, namely, that there is no difference, subjectively, between a "square" and a "short rectangle."

Experiment II.—This experiment is a repeat of Cond. A but with 10 new *S*s. This was done since Cond. A is a necessary test and the data on the first 3 *S*s were not recorded. Six males and 4 females were tested for visual acuity with the orthorater and had better than 20/20 vision. Condition A differs from B and C in that *S* had three categories: tall rectangle, square, and short rectangle. Table 3 gives the T^2 values in testing for significance between the mean for the subjective square and the mean for the subjective short rectangle categories. It was assumed that the tall rectangle would differ significantly from the other categories. It can be seen in Table 3 that only 4 *S*s had two distinct categories and 6 had only one. Of these 4, 3 were significant at the .01 level and the fourth at the .05 level. When averaged, the 10 *S*s, as a group, did not discriminate significantly (less than .05) between the square and the short rectangle categories.

TABLE 3

T^2 VALUES IN TESTING FOR A DIFFERENCE BETWEEN THE MEAN FOR THE SQUARE CATEGORY AND THE MEAN FOR THE SHORT RECTANGLE CATEGORY

*S*s	T^2	*df*
1	2.50	2, 60
2	2.17	2, 42
3	10.76**	2, 47
4	3.24	2, 50
5	4.04*	2, 67
6	2.95	2, 35
7	9.91**	2, 57
8	2.19	2, 51
9	12.07**	2, 33
10	1.79	2, 56
M value	3.66	2, 50

Note.—The *S*s had three categories: SQ, TR, and SR.

* $p < .05$.

** $p < .01$.

For the 4 *S*s who did discriminate, 2 (S_3 and S_9) placed the square category between the two rectangle categories. This was determined by finding the h/w for the mean of the square category. The h/w for the mean square lay between the h/w for the two mean rectangles. The other 2 *S*s (S_5 and S_7) placed the mean for the square category beyond the mean for the short rectangle. That is to say, the h/w for the mean of their square category was less than the h/w for the mean of their short rectangle category. Thus, only 2 *S*s out of 10 performed as the subjective-standard hypothesis would predict. When considered individually, or as a group, these *S*s did not have three distinct categories although they had three different responses.

DISCUSSION

These results imply that the subjective-standard hypothesis, with its notion of an internal standard or memory category, is inadequate in accounting for these data. For *S*s in both experiments there, was, in general, no difference between the "square" category and the "short rectangle" category. The conclusion, then, is obvious: "square," in this study, is a verbal label only and has no subjective referent. The data demonstrate that no difference exists between the mean for the stimuli assigned to the square category by *S* and those assigned to the short rectangle category for different judgments and at different times. Table 3 shows the same thing within the same judgmental situation. The *S* classified the stimuli the same way, i.e., used different labels for the same stimuli in different judgmental conditions.

This seems to be a paradox but the fallacy here lies in equating responses as labels, with perceptual (or memory) categories. No one can deny the stability of certain responses as labels; certain words, phrases, etc. have high probabilities of occurrences. This is not sufficient evidence for the existence of referents for these labels. To make this point clear, consider color-blind individuals who can, and do, use certain words, e.g., "blue," "green," etc. and yet have no subjective referent for these words. Thus, from an ability to use a word, it cannot be inferred that *S* has a referent. For an opposite position, see Vernon (1955) and Binder (1955).

We now turn and look at the notion of memory categories more closely. Different writers stress different aspects but there is a logic which is common to all. First, there is the notion of a stimulus trace somewhere in the nervous system (Wulf, 1922), or a pattern of neural firing in the nervous system (Hebb, 1949). In either case the presented stimulus object produces neural excitation and, with repeated presentations, eventually wears a path in the cortex. This is similar to Hebb's (1949) "cell assembly." Once the pattern of firing has occurred enough times to create a pathway, experiences are correlated with paths. This is a sort of physiological matching in that it generated a particular pattern of neural firing which is similar to some previously generated path. Thus triangles are perceived as triangles because they generate a particular pattern of excitation in the cortex; all triangles generate the same pattern otherwise they would not be experienced as triangles.

On a behavioral level this kind of logic leads to a "library" theory of memory. The input enters the "library" and is to be classified; it is cross-indexed according to its size, hue, perimeter, number of angles, etc. and then assigned to that "shelf" or "stack" which has objects most similar to it. This approach also allows an apparent categorical, or absolute, judgment to be a comparative one.

Both the physiological and behavioral positions maintain that the subjective-standard hypothesis must be in line with the facts of perception. Nevertheless, the data reported here constitute a strong counterargument. The position taken in this paper is that the perceptual system is a process which uses criteria or a set of principles in classifying experience and which are not contingent on memory. The difference between this position and that taken by Eriksen and Hake (1957) and James (1963) comes about in specifying the set of principles (criteria) used in making perceptual judgments. Everyone agrees that criteria exist and are used when individuals make perceptual judgments. The subjective-standard position maintains that there is one basic principle, namely, the perceptual system compares the input with a stored category and assigns the input to the category which it most closely resembles. On the other hand, the coherence criterion notion says that there exists a set of principles that determines the category to which the input will be assigned, and that this set of principles is determined by the interrelationships among the components of the stimulus object.

The purpose of this study was to demonstrate that the subjective standard, in its naive form, is incapable of dealing with certain phenomena. In order for the subjective-standard hypothesis to account for the data presented here, a change must be made. The notion of a perceptual category being similar to an image cannot be kept, and most importantly, a category can be defined as a judgmental rule, i.e., as a rule of behavior which is appropriate to the present situation.

REFERENCES

Anderson, T. W. *An introduction to multivariate statistical analysis.* New York: Wiley, 1958.

Binder, A. A statistical model for the process of visual recognition. *Psychol. Rev.*, 1955, **62,** 119–129.

Eriksen, C. W., & Hake, H. W. Anchor effects in absolute judgments. *J. exp. Psychol.*, 1957, **53,** 132–139.

Hake, H. W., Rodwan, A. S., & Weintraub, D. J. Noise improvement in vision. In K. Hammond (Ed.), *Egon Brunswik's psychology.* New York: McGraw-Hill, in press.

Hebb, D. O. *The organization of behavior.* Wiley: New York, 1949.

James, W. *Psychology, briefer course.* (2nd ed.) Greenwich, Conn.: Fawcett Publications, 1963.

Rodwan, A. S. An empirical validation of the concept of coherence. *J. exp. Psychol.*, 1964, **68,** 167–170.

Rodwan, A. S., & Hake, H. W. The linear discriminant function as a model for perception. *Amer. J. Psychol.*, 1964, **77,** 380–392.

Swets, J. A., Tanner, W. P., Jr., & Birdsall, T. G. Decision processes in perception. *Psychol. Rev.*, 1961, **68,** 301–340.

Vernon, M. D. The functions of schemata in perceiving. *Psychol. Rev.*, 1955, **62,** 180–192.

Wulf, F. Beitrage zur psychologie der Gestalt: VI Uber die Veranderung von Vorstellungen (Gedachtniss und Gestalt). *Psychol. Forsch.*, 1922, **1,** 333–373.

(Received April 8, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 1, 63–74

IDENTIFICATION AND DISCRIMINATION FUNCTIONS FOR A VISUAL CONTINUUM AND THEIR RELATION TO THE MOTOR THEORY OF SPEECH PERCEPTION [1]

D. V. CROSS, H. L. LANE, AND W. C. SHEPPARD

Center for Research on Language and Language Behavior, University of Michigan

The motor theory of speech perception maintains that "articulatory movements and their sensory effects mediate between the acoustic stimulus and the event we call perception [Liberman, 1957, p. 122]." The inference of mediating articulation is based on certain properties of identification and discrimination functions for synthetic speech continua, namely, categorical identification and enhancement of discrimination at the boundary between identification classes. Identification and discrimination functions with these properties also were obtained in the present experiment, although a nonspeech (visual) continuum was employed. It is concluded that the postulation of a special perceptual mechanism for speech perception is not warranted.

Advocates of the motor theory of speech perception (Liberman, Cooper, Harris, & MacNeilage, 1963) contend that "speech sounds are perceived by reference to the articulatory movements that produce them, and this articulatory reference is important for the distinctiveness of speech as perceived [p. 6]." According to this view, the sounds of speech are absolutely and quickly identifiable because continuous variations in the acoustic stimulus cue phonemic distinctions that are categorical in nature for most speech continua. The theory maintains (Liberman et al., 1963) that the categorical perception of speech arises from a learned connection between speech sounds and the articulations that generate them, and that "in time, these articulatory movements . . . come to mediate between the incoming acoustic stimulus and its ultimate perception [p. 4]." Stimuli from the same phoneme class evoke the same mediating response and acquire similarity, whereas stimuli from different phoneme classes evoke different mediating responses and gain the added distinctiveness of their different proprioceptive returns. Thus, speech perception is distinctive because it is categorical, and it is categorical because the mediating articulations are categorically different.

Proponents of the motor theory of speech perception find empirical support for their hypothesis in experiments involving typically two dependent variables: identification functions and discrimination functions for synthetic speech continua. The two experiments reported in this article examine identification and discrimination functions, respectively, for a nonspeech (visual) continuum. Our purpose is to determine whether the properties of speech perception on which the motor theory is founded are obtained exclusively with speech stimuli or whether, more generally, these properties may be obtained also with stimuli which neither are generated by their identification responses

---

[1] This research was supported in part by the Language Development Section, United States Office of Education (Contract SAE-9265) and the National Science Foundation (Grant No. GB-552).

nor have a long history of connection with them.

EXPERIMENT I: IDENTIFICATION FUNCTIONS FOR A VISUAL CONTINUUM

The first kind of evidence adduced in support of the motor theory of speech perception comes from studies attempting to show that phoneme perception is discontinuous (categorical) or continuous depending on whether the appropriate articulations are discontinuous or continuous. For example, an important acoustic cue for the perception of the voiced stops /b/, /d/, and /g/ is the direction and extent of the second-formant transition. When a series of synthetic speech sounds, varied in small steps along this continuum, is presented to listeners who are asked to identify the sounds as /b/, /d/, or /g/, their responses characteristically divide the continuum into three sharply defined phoneme categories (Liberman, Harris, Hoffman, & Griffith, 1957). The shifts from one response or phoneme label to another are abrupt. On the other hand, when listeners are asked to identify synthetic vowels, which vary in small steps along an appropriate acoustic continuum, their responses do not divide the continuum into sharply defined phonemic categories; instead, the shift from one category to another is gradual (Fry, Abramson, Eimas, & Liberman, 1962). The voiced stops in the first example are produced by different articulations; "/b/ is produced by a movement of the lips and /d/ by a movement of the tongue. . . . There are no intermediate articulations between /b/ and /d/ [Liberman et al., 1963, p. 4]." In the production of vowel sounds, however, the articulators can move continuously from the resonance positions characterizing one vowel phoneme to those of the next. Thus, the theory maintains that the perception of most consonants is discontinuous because they covertly evoke discontinuous (i.e., discretely different) articulations, whose sensory feedback controls the overt identifications. Accordingly, the perception of vowels is continuous because the mediating articulatory movements are continuous.

The statement that perception is categorical for consonants and continuous for vowels is argumentative rather than factual. It is based on comparisons of identification and discrimination functions relative to *different* synthetic speech continua; since such comparisons are not dimensionally homogeneous, the continuous-discontinuous distinction is arbitrary (Lane, 1965). With appropriate conditioning techniques, however, it is possible to transfer stimulus control of the two kinds of phonemic contrasts to a single nonspeech continuum. Direct comparison of consonant and vowel identification functions would then be meaningful, since they would be relative to the same coordinate system. The purpose of the two studies comprising Exp. I is, therefore, twofold: to determine whether categorical perception is obtained with a nonspeech visual continuum and, if so, whether categorical perception is related to the topography of the identification responses, as the motor theorists would have it. In these studies, a consonant contrast (/b/, /g/) or a vowel contrast (/ ʌ /, /i/) was conditioned to a pair of visual stimuli. Measures of response frequency and latency were then obtained at intermediate stimulus values, using procedures that were as comparable as possible to those employed by Liberman and his collaborators.

Method

Experiment Ia

Subjects.—Twenty paid volunteers, 11 male and 9 female undergraduates, served individually in 1-hr. sessions.

Apparatus.—The *S* wore a dynamic throat microphone (Danavox) while seated, in a sound-insulated chamber, in front of a projection screen and a coin dispenser. A slide projector (Leitz VIII S), located behind *S* and outside of the chamber, projected the visual stimuli through a window and onto the screen. A vocal response to the stimulus triggered a voice-operated relay, which closed an electrically operated shutter (rise-decay time, 40 msec.), advanced the projector to the next slide, and opened the shutter 2 sec. later. The latency of the response was measured in milliseconds on a frequency counter (Hewlett-Packard 523D), which started when the shutter was signaled to open and stopped when it was signaled to close. The form of the response was monitored and recorded by *E*.

Stimuli.—Nine discs, 3 in. in diameter, were cut out of a piece of flat black art paper. A sector, symmetrical about the radius at 360°, was cut out of each disc; the sectors that were deleted ranged from 30° to 62° in steps of 4°. The nine sectors remaining were pasted on pieces of white paper and transparencies were prepared with a Polaroid-Land copy-maker (Model 208) and film (Type 146-L). The stimuli had a diameter of 10 in. when projected on the screen, 5 in. above eye level and 24 in. in front of *S*. The stimulus field had a luminance of approximately 10^{-3} foot-lamberts (ftL.). Eight irrelevant stimuli, also prepared and presented in this manner, were used in identification training. These stimuli consisted of various geometrical patterns, based on variants of the cross, rhombic, polygon, and solar designs.

Procedure.—The 20 *S*s were assigned randomly to two groups of 10 each. Those in Group V (vowel contrast) were conditioned to identify the extreme stimuli in the series, No. 1 and 9, as /bʌb/ and /bib/, respectively; *S*s in Group C (consonant contrast) were conditioned to call them /bʌb/ and /gʌg/. Thus, the identification responses for Group V were drawn from an articulatory continuum, while those for the other group (C) were "essentially discontinuous."[2]

Identification training. These instructions were read to *S:* (The alternative words, in parentheses, were used with Group C.)

> On this screen will appear various symbols. Some are /bʌb/ symbols and some are /bib/ (/gʌg/) symbols; still other symbols are irrelevant. Your task is to learn to identify the /bʌb/ symbols and the /bib/ (/gʌg/) symbols when they appear, by saying /bʌb/ or /bib/ (/gʌg/). Each time a /bʌb/ or /bib/ (/gʌg/) symbol is presented and you correctly identify it, you will get a nickel, which you may keep. If you respond incorrectly to one of the /bʌb/ or /bib/ (/gʌg/) symbols, or if you respond to any of the irrelevant symbols, no nickel will be dispensed.

Stimuli 1 and 9 were presented 16 times each and the eight irrelevant stimuli five times each, all in a randomized order. Occurrences of the /bʌb/ response within 3 sec. after the presentation of Stimulus 1 were reinforced, while those of the /bib/ (or /gʌg/) response were not. The relations were reversed for Stimulus 9. The stimulus duration equaled the response latency or 3 sec., whichever occurred first. All *S*s met the training criterion of 10 consecutive reinforced responses within the series of 72 stimuli.

Identification testing. These instructions were read to *S:*

> In this part of the experiment we will present only /bʌb/ and /bib/ (/gʌg/) symbols. Your task is to indicate whether the symbol presented each time is an example of a /bʌb/ symbol or /bib/ (/gʌg/) symbol. We are going to disconnect the nickel dispenser, so you will not know whether your response is correct, but I will be keeping score and I will tell you how well you did when we are through. Be sure to respond to each presentation—guess if necessary.

The nine stimuli were presented, 10 times each, in a randomized order. The stimulus duration equaled the response latency of *S*, and the interstimulus interval was 2 sec.

Experiment Ib

The procedure of Exp. Ia was replicated on 7 consecutive days with each of two female and two male paid *S*s. The same stimuli and the same responses were used: two of the *S*s were assigned the vowel contrast and two were assigned the consonant contrast.

Results

Identification probability functions and their associated latency functions

---

[2] Vide Liberman et al., 1957.

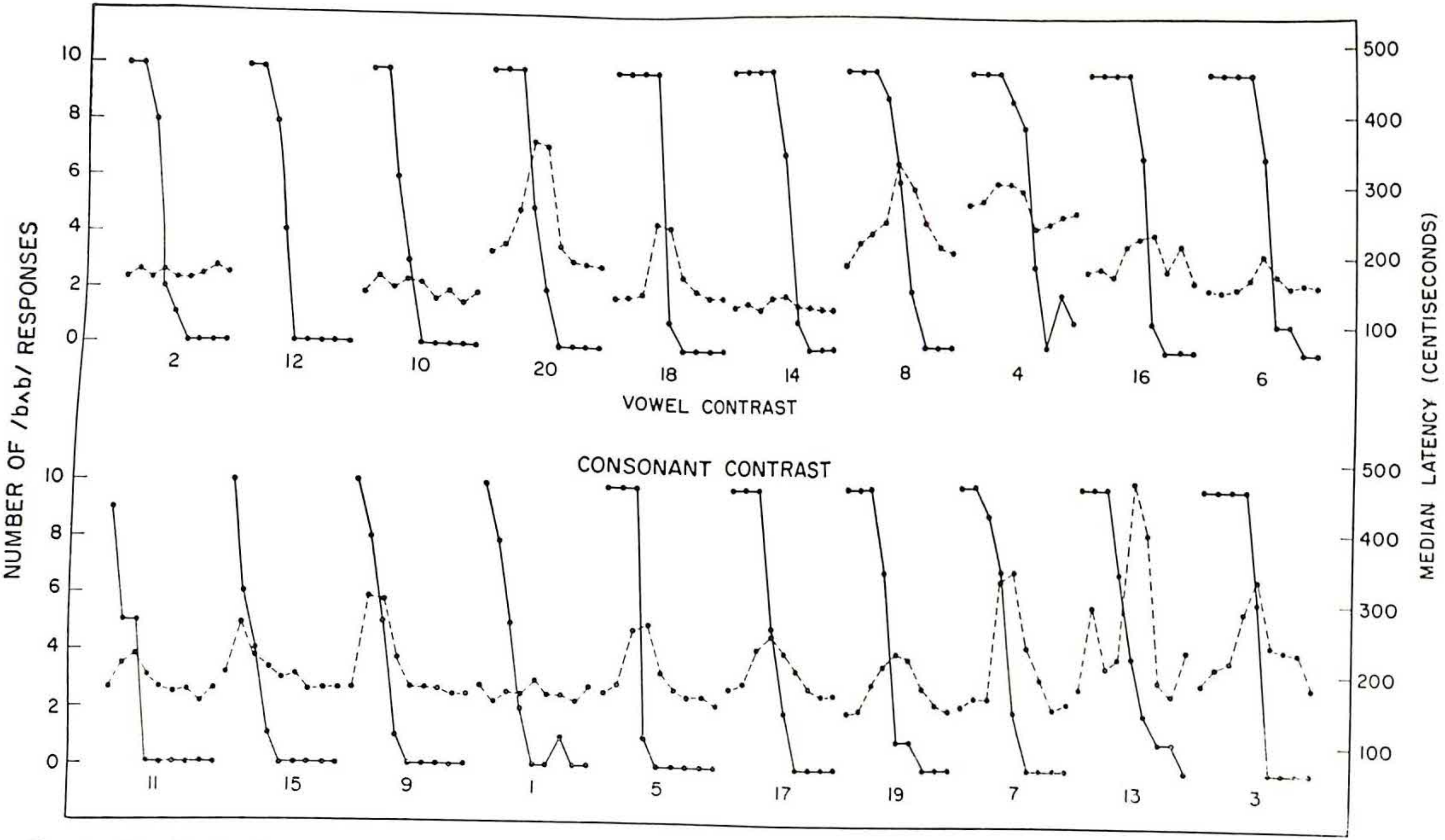


FIG. 1. Identification frequency and latency functions for individual *S*s in Exp. I. (The points connected by continuous lines represent the number of /b∧b/ responses evoked by the 10 presentations of each stimulus. The *S*s were numbered relative to the order in which they served in the experiment but their identification functions are ordered, from left to right, roughly on the basis of where each gradient and its complement intersected on the stimulus continuum. The points connected by broken lines represent the median latencies of all responses emitted to the corresponding stimulus. Latencies were not measured for *S* 12.)

are shown in Fig. 1 for individual *S*s from Exp. Ia. The identification probability functions show the number of times each stimulus was identified as a /b ʌ b/ symbol in the 10 presentations of that stimulus. (The /bib/ or /g ʌ g/ identification functions are complementary to those presented and are not shown.) The identification functions of all *S*s are flat at one or both extremes and over an extended range of the stimulus continuum. Thus, they may be described as partitioning the stimulus set into two classes; the stimuli in each class are equally effective in evoking a given identification response.

Response latencies, measured as the time elapsing from the onset of the stimulus to its termination by either response, are represented in Fig. 1 (dashed lines) by their median value at each stimulus. Although each determination is based on only 10 responses, the latency functions are orderly for most *S*s and show a close relation to the corresponding identification function: (*a*) maximum latencies occur at or near the point where the pair of identification responses are equiprobable; (*b*) latencies are approximately equal over the stimulus ranges corresponding to plateaus in the identification function.

The results obtained in Exp. Ib, with measurements repeated on 7 consecutive days, are in accord with those reported above. Identification func-

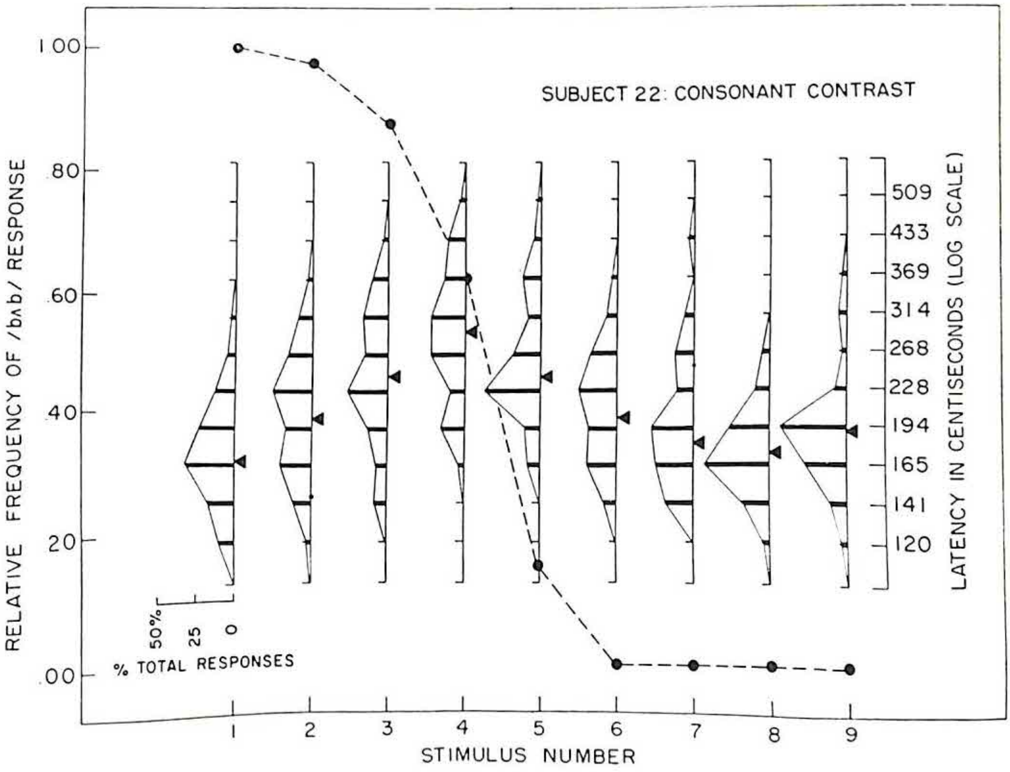


FIG. 2. Identification probability and latency for one *S* from Exp. Ib (consonant contrast). (The relative frequency of the /b ʌ b/ response, in 60 presentations of the corresponding stimulus on each of 6 consecutive days, is shown by the dashed lines. Frequency distributions of the latencies of the /b ʌ b/ and /bib/ responses combined are shown at each stimulus value; the median latency is represented by the filled triangle.)

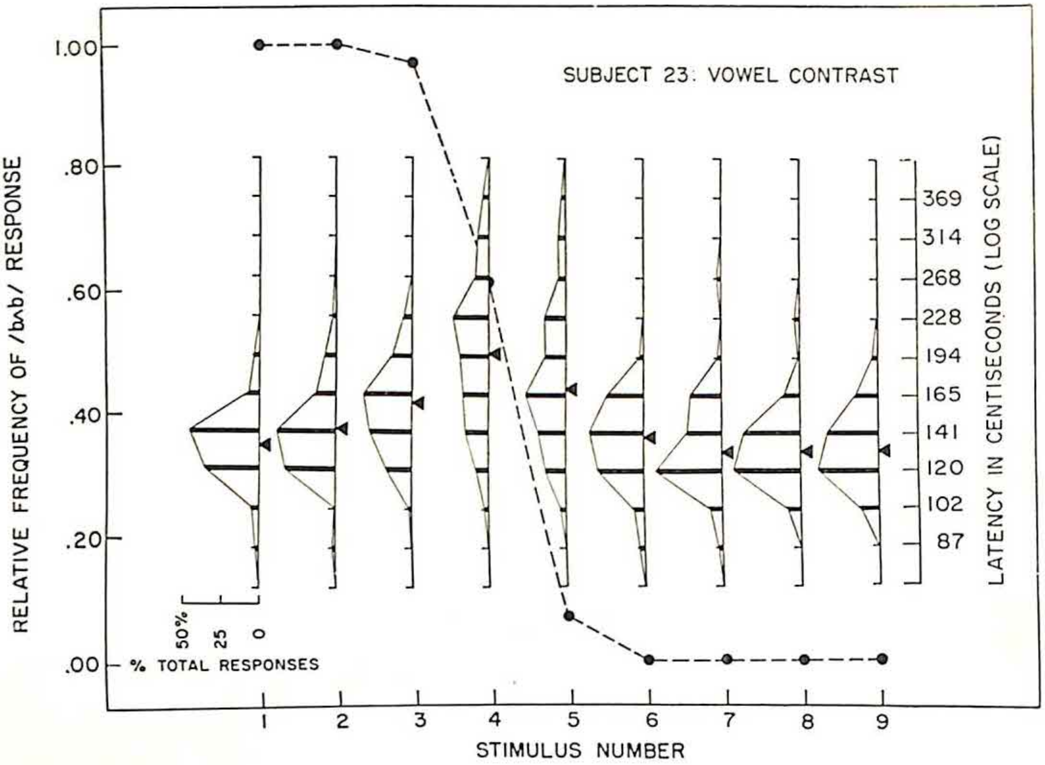


FIG. 3. Identification frequency and latency for one *S* from Exp Ib (vowel contrast). (The relative frequency of the /b∧b/ response, in 60 presentations of the corresponding stimulus on each of 6 consecutive days, is shown by the dashed lines. Frequency distributions of the latencies of the /b∧b/ and /g∧g/ responses combined are shown at each stimulus value; the median latency is represented by the filled triangle.)

tions based on the last 6 days of testing, are shown in Fig. 2 and 3 for one *S* from each experimental group. Each data point represents the proportion of times that the corresponding stimulus was identified as a /b∧b/ symbol in 60 presentations. Response latencies are represented by histograms showing the distribution of latencies with respect to 10 equal class intervals on a logarithmic scale. The approximate midpoint of each class interval is indicated on the latency scales in Fig. 2 and 3. The median latency for each distribution is indicated by a filled triangle. The identification functions for these two *S*s are nearly identical and their latency distributions reveal the same general functional relation to response probability. Response latencies for *S* 22 are longer overall than those for *S* 23 and show greater variability, but the latencies of both *S*s are distributed over their minimum values when the probability of one of the identification responses is either zero or unity, and the latencies are distributed over their maximum values when the probability of the responses is near .50.

EXPERIMENT II: DISCRIMINATION FUNCTIONS FOR A VISUAL CONTINUUM

The second kind of evidence often adduced in support of the motor theory of speech perception comes from studies which show that a listener discriminates more accurately between sounds that lie on opposite sides of the phoneme (identification) boundary than he does between sounds (separated by the same number of physical units) drawn from within one phoneme category (Liberman, Harris, Eimas, Lisker, & Bastian, 1961; Liberman et al., 1957; Liberman, Harris, Kinney, & Lane, 1961). The motor theory attributes this outcome to the role of mediating responses in speech perception: the several stimuli within a phoneme class all evoke the same covert response and are associated, therefore, with identical proprioceptive information; stimuli drawn from opposite sides of the phoneme boundary evoke different mediating responses and are associated, therefore, with distinctively different proprioceptive returns: ". . . The perceived difference between two relatively similar external stimuli could be increased if we could attach to those stimuli two very different mediating responses and hence gain the added distinctiveness of their very different proprioceptive returns [Liberman, 1957, p. 123]."

The purpose of Exp. II is to determine whether an *O* who has learned to identify categorically stimuli drawn from a nonspeech continuum, discriminates more accurately between stimuli that lie on opposite sides of the identification boundary than he does between stimuli (separated by the same number of physical units) drawn from the same identification category.[3]

Method

Subjects.—Four male undergraduate students served individually in sessions lasting approximately 1 hr.; none had participated in Exp. I.

Stimuli and apparatus.—The stimuli numbered 4, 5, 6, and 7 in Exp. I were employed; they had sectors of 42, 46, 50, and 54°, respectively, deleted from their rims. The stimuli were presented with the same apparatus employed in the preceding studies.

Procedure.—The experimental sessions consisted of three phases in the following order: (*a*) identification training, (*b*) identification testing, (*c*) discrimination testing.

Identification training. These instructions were read to *S:*

> On this screen will appear various symbols which will differ only with respect to the size of the wedge cut out of them. Your task is to identify each symbol as it is presented by saying /b∧b/ or /g∧g/—whichever you think is appropriate. Each time a symbol is presented and you correctly identify it, you will get a penny. If you respond incorrectly, a penny will not be dispensed. The faster you learn to correctly identify these symbols, the more money you will earn.

The /b∧b/ response was reinforced promptly in the presence of Stimulus 4 or Stimulus 5, the /g∧g/ response in the presence of Stimulus 6 or Stimulus 7. Each stimulus presentation lasted 300 msec. The four stimuli were presented in random order until 50 successive correct responses terminated the phase.

Identification testing. These instructions were read to *S:*

> In this part of the experiment we will present the same symbols. As before, your task is to identify each as a /b∧b/ symbol or a /g∧g/ symbol. However, we are going to disconnect the coin dispenser, so you will not know whether or not each response is correct. I will keep score and tell you how well you did when we are through. Be sure to respond to each presentation—guess if you have to

The sequence of 96 stimulus presentations was prepared so that all permutations of the four stimuli appeared in random order.

Discrimination testing. These instructions were read to *S:*

> In this part of the experiment we want to find out how well you can discriminate /b∧b/ and /g∧g/ symbols. We will present three symbols in a sequence. The third symbol will always be identical to either the first or the second. When the third symbol is presented, you are to say "first" if you think it is like the first symbol or to respond "second" if you think it is like the second symbol. Respond only to the third symbol—guess if you have to, but be sure to respond.

---

[3] We wish to thank George Lombard for his assistance in conducting this experiment.

The ABX procedure employed by Liberman and his colleagues was adopted here to insure that our results would be comparable to theirs in this regard. Only one-step comparisons were employed; each pair of adjacent stimuli was presented 36 times in all possible triadic arrangements of the form ABA, ABB, BAA, and BAB. The stimulus duration was 300 msec., the interval between stimuli 5 sec., and that between triads 15 sec.

Results

The training criterion was achieved rapidly by all four *S*s (within 80 to 120 stimulus presentations). The results of the identification and discrimination tests are shown for each *S* separately in Fig. 4. The identification functions (filled circles, solid lines) show the relative frequency of the /b∧b/ response in the 24 presentations of each stimulus. The discrimination functions (open circles, dashed lines) show the relative frequency of correct discriminations in the 36 presentations of each stimulus pair. For all *S*s, the stimuli that evoked different identification responses were more accurately discriminated than those that evoked the same identification response.

DISCUSSION

The findings reported above demonstrate that stimulus control of phonemic contrasts is readily transferred to a nonspeech continuum and that the prop-

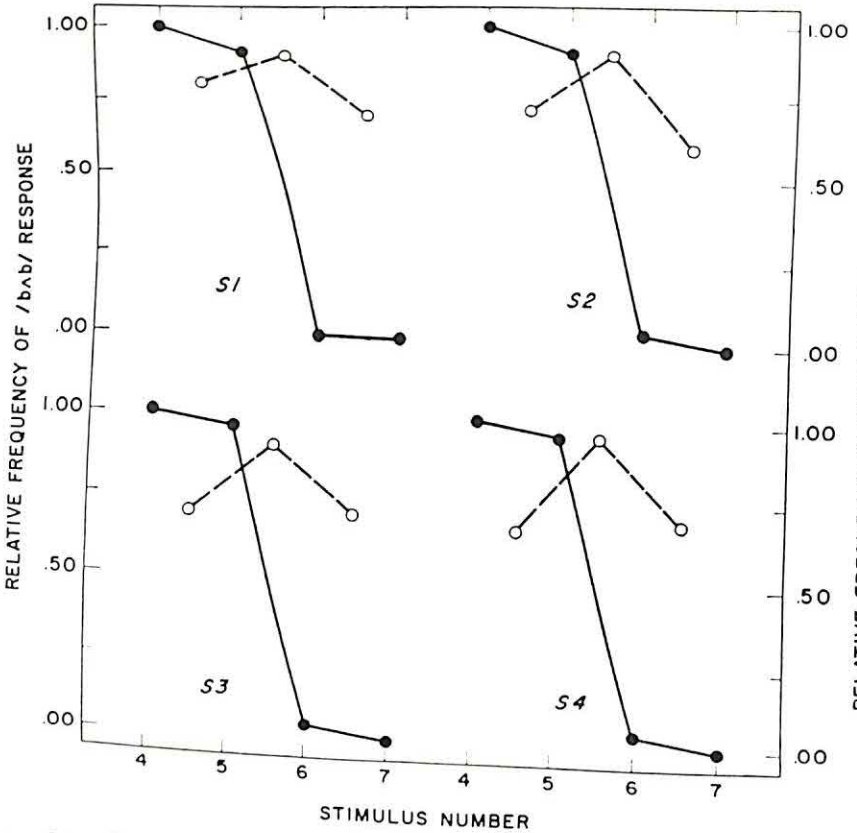


FIG. 4. Identification and discrimination of visual stimuli by individual *S*s. (Each point on the identification functions—filled circles, solid lines—represents the relative frequency of the /b∧b/ response in 24 presentations of the corresponding stimulus. Each point on the discrimination functions—open circles, dashed lines—represents the relative frequency of correct discriminations in 36 presentations of each pair of stimuli.)

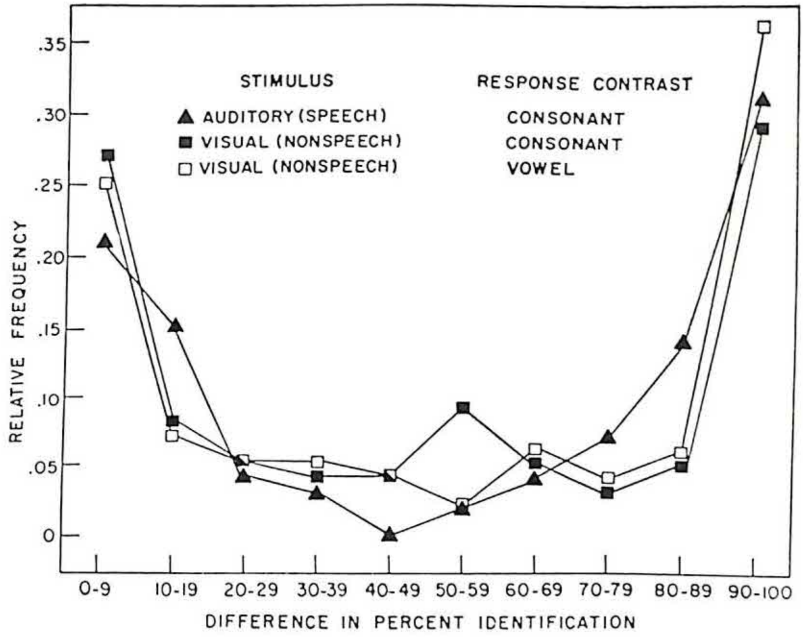


FIG. 5. A comparison of the degree of categorical perception obtained with identification of synthetic speech stimuli and identification of visual forms. (The triangles show the transformed labeling gradients for the phonemic contrast /do/–/to/, reported in Liberman, Harris, Kinney, & Lane, 1961. The open and filled squares represent a transformation of the labeling gradients shown on the top and bottom of Fig. 1, respectively, for identification of visual forms.)

erties of stimulus identification and discrimination formerly attributed exclusively to speech perception are preserved in this transfer.

Identification functions for speech and nonspeech continua.—The identification functions obtained in this study show response probabilities that are constant at zero or unity over wide ranges of stimulus variation. In Fig. 5, the results obtained in Exp. I are compared with those obtained with synthetic speech stimuli (/do/, /to/) in a study by Liberman, Harris, Kinney, and Lane (1961). Since different stimulus continua were involved in these two studies, a direct comparison of the respective identification functions would be inappropriate. Therefore, the comparison is based on the following procedure. For each *S*, each stimulus was compared to every other in the percentage of /bʌb/ responses it evoked, in the case of the present study, or in the percentage of /do/ responses evoked, in the case of the synthetic speech study. The difference in percentage was computed for each comparable pair. A frequency distribution of these percentage differences was prepared, with class intervals of 10%, converted to relative frequencies and plotted in Fig. 5. If the individual gradients were step functions, the limiting case of "categorical perception," then all the percent differences would be either 0 or 100, and the frequency distribution would pile up in the extreme classes, leaving intermediate classes empty. On the other hand, if perception was continuous—the case where response probability is a strictly monotonic function of stimulus change—then the above frequency distribution would be more nearly rectangular in shape. Figure 5 shows that the distributions for Groups V and C in the visual-identification task do not differ systematically from each other nor do they differ appreciably from those for *Ss* performing a speech-identification task, although the set of speech-identification gradients selected for comparison is the most "categorical" of those reported in the literature.

The response-latency functions obtained in this study, with both vowel-contrast and consonant-contrast labeling responses, reach a maximum at the "boundary" of their corresponding identification functions and a minimum within the plateaus of those functions. A closer look at this relation between measures of response probability and latency is provided in Fig. 2 and 3. Over the range of stimuli which are equally effective in evoking a given identification response (with either unit or zero probability) the latency distributions are highly uniform: the different distributions have approximately the same mean and variance. The latencies of responses evoked by boundary stimuli are more variable and are distributed over greater values. These findings for response latency in identification of nonspeech visual forms accord with those reported for identification of synthetic speech stimuli (although the latter findings were interpreted as favoring a mediation hypothesis): "For both stops and vowels, voice and button-press reaction times were lower near phoneme centers than in the vicinity of phoneme boundaries [Studdert-Kennedy, Liberman, & Stevens, 1963, p. 1009]. Peaks in latency functions at boundary stimuli have been reported by several investigators and for diverse nonspeech continua: size of angles (Cartwright, 1941), brightness (Kellogg, 1931), vertical height (LaBerge, 1961), pure tone and noise intensity (Cross & Lane, 1962)—to name only a few. These findings combine to support the following generalization: the average latency of two concurrent discriminative operants is maximal when the responses are equiprobable. This generalization applies to nonspeech as well as to speech continua; if it is interpreted as implying that perception is categorical, it is, nevertheless, not peculiar to speech perception.

Advocates of the motor theory of speech perception have cited the properties of the identification probability and latency functions described above as partial evidence that perception is categorical and is mediated by covert articulation (or its neural correlates). If we conclude, therefore, that 20 min. of discrimination training sufficed to transfer the mediation processes of speech perception to a visual continuum, then we must ask why the consonant contrasts did not yield more categorical gradients of identification and latency then the vowel contrasts: if categorical perception is effected by discretely different mediating responses, then the consonant contrasts should have been much sharper than the vowel contrasts. If it is maintained, on the other hand, that the mediation hypothesis is applicable only to the perception of speech continua and is predicated on extensive verbal conditioning (as the motor theorists would have it), then we must ask why the identification and latency functions for visual symbols so closely resemble those for speech signals. Since visual forms and synthetic speech stimuli yield equivalently categorical identification functions and peaked latency functions, these dependent variables provide no evidence which would warrant the inference of a mechanism uniquely mediating the perception of speech.

Discrimination functions for speech and nonspeech continua.—In several studies involving synthetic speech stimuli it has been found that when two stimuli evoke *different* identification responses they are more readily discriminated than are stimuli (separated by the same number of physical units) that evoke the *same* identification response. It has been claimed that this relation between speech identification and discrimination functions is evidence for a motor theory of speech perception. The rationale is spelled out by Liberman, Harris, Eimas, Lisker, and Bastian (1961):

> We believe that in the course of his long experience with language a speaker (and listener) learns to connect speech sounds with their appropriate articulations. In time, these articulatory movements and their sensory feedback (or, more likely, the corresponding neurological processes) become part of the perceiving process, mediating between the

> acoustic stimulus and its ultimate perception. When significant acoustic cues that occupy different positions along a single continuum are produced by essentially discontinuous articulations . . . the perception becomes discontinuous (i.e., categorical), and discrimination peaks develop at the phoneme boundary [p. 177].[4]

The findings in Exp. II of the present study (Fig. 4) show that the enhancement of discrimination at identification boundaries is not peculiar to stimuli that have a long history of connection with the responses that generate them. In this study a relatively brief training period sufficed to bring two discriminative operants (vocal) under the control of a visual continuum that had no prior connection with either response.

Lane (1965) has reported a comparable experiment in which the discrimination of certain nonspeech auditory patterns (the inverted /do/-/to/ sounds employed as control stimuli by Liberman, Harris, Kinney, & Lane, 1961) was measured both before and after discrimination training with the two extreme values of the stimulus parameter. Lane found chance levels of discrimination in all ABX comparisons before identification training. Following this training, ABX discrimination functions showed peaks at the identification boundary in one-, two-, and three-step comparisons for all *S*s. The peaks were comparable to those obtained by Liberman, Harris, Kinney, and Lane (1961) in discrimination functions for /do/-/to/ speech stimuli. Lane (1965) concluded that since the correlation between identification probability and discrimination accuracy is observed for nonspeech as well as for speech stimuli, the correlation per se does not seem to warrant the postulation of a special perceptual mechanism for the discrimination of speech stimuli.

An alternative formulation.—It appears that the observed correlation between identification probability and discrimination accuracy is not conditional on whether or not the stimuli are perceived as speech sounds but, more broadly, on whether prior discrimination training has enabled *S* to absolutely identify certain elements of the stimulus set. The findings of an experiment by Chistovich, Klaas, and Alekin (1961) support this inference. These investigators presented rapid sequences of three tones (300, 1,200, 3,500 cps) to 14 *S*s with instructions, in one experiment, to identify the order in which the low-, middle-, and high-pitch tones appeared in each sequence by writing LMH, HLM, etc. In a second experiment, *S* was trained to identify the tones by emitting vocal responses /u/, /a/, and /i/, to the 300-, 1,200-, and 3,500-cps tones, respectively, and then to identify the order in which the tones appeared in the test sequences by writing UAI, IUA, etc. The stimuli were chosen to match the quality of the corresponding vowels; they were varied in their order within triads and in their duration. There was a highly significant improvement in the discriminability of tone sequences when *S* identified them as vowels.

Since the stimuli in an ABX triad are presented in temporal succession, specific properties of a component stimulus may lose control of discriminative responding by the time the next stimulus comes along. However, if an *O* has been trained to absolutely identify certain of these stimuli, the cues provided by this simple binary response may provide a basis for discrimination long after the relevant stimuli are gone. If this conjecture is sound, it explains why both speech and nonspeech stimuli that evoke different identification responses are discriminated in ABX tasks more readily than stimuli that evoke the same identification response. The *O* is discriminating certain gross properties of his own behavior rather than the specific properties of the stimuli.

The findings presented in this article when taken together with those reported in a review of the motor theory of speech perception by Lane (1965) make it clear that the relations among identification probability, latency, topography, and discrimination accuracy on which the motor theory is based are not at all peculiar to speech perception but are, more broadly, the result of a rather general paradigm for discrimination training and testing.

---

[4] "Given the discontinuous articulation [of /b-d-g/], we should expect, in spite of the continuous nature of the acoustic variations, that the perception would be discontinuous (i.e., categorical), and that a discrimination peak would appear at the phoneme boundary [Liberman et al., 1963, p. 4]."

REFERENCES

CARTWRIGHT, D. Relation of decision-time to the categories of response. *Amer. J. Psychol.*, 1941, **54**, 174–196.

CHISTOVICH, L. A., KLAAS, YU. A., & ALEKIN, R. O. The role of imitation in discrimination of sound series. *Vop. Psikhol.*, 1961, **5**, 173–182.

CROSS, D. V., & LANE, H. L. On the discriminative control of concurrent responses: The relations among response frequency, latency, and topography in auditory generalization. *J. exp. Anal. Behav.*, 1962, **5**, 487–496.

FRY, D. B., ABRAMSON, A. S., EIMAS, P. D., & LIBERMAN, A. M. The identification and discrimination of synthetic vowels. *Lang. Speech*, 1962, **5**, 171–189.

KELLOGG, W. N. The time of judgment in psychometric measures. *Amer. J. Psychol.*, 1931, **43**, 65–86.

LABERGE, D. Generalization gradients in a discrimination situation. *J. exp. Psychol.*, 1961, **62**, 88–94.

LANE, H. L. Motor theory of speech perception: A critical review. *Psychol. Rev.*, 1965, **72**, in press.

LIBERMAN, A. M. Some results of research in speech perception. *J. Acoust. Soc. Amer.*, 1957, **29**, 117–123.

LIBERMAN, A. M., COOPER, F. S., HARRIS, K. S., & MACNEILAGE, P. F. A motor theory of speech perception. In *Proceedings of the Speech Communication Seminar.* Stockholm: Royal Institute of Technology, 1963.

LIBERMAN, A. M., HARRIS, K. S., EIMAS, P., LISKER, L., & BASTIAN, J. An effect of learning on speech perception: The discrimination of durations of silence with and without phonemic significance. *Lang. Speech*, 1961, **4**, 175–195.

LIBERMAN, A. M., HARRIS, K. S., HOFFMAN, H. S., & GRIFFITH, B. C. The discrimination of speech sounds within and across phoneme boundaries. *J. exp. Psychol.*, 1957, **54**, 358–368.

LIBERMAN, A. M., HARRIS, K. S., KINNEY, J., & LANE, H. L. The discrimination of relative onset-time of the components of certain speech and nonspeech patterns. *J. exp. Psychol.*, 1961, **61**, 379–388.

STUDDERT-KENNEDY, M., LIBERMAN, A. M., & STEVENS, K. N. Reaction time to synthetic stop consonants and vowels at phoneme centers and at phoneme boundaries. *J. Acoust. Soc. Amer.*, 1963, **35**, 1009.

(Received April 9, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 1, 75–78

SHORT-TERM RETENTION OF VERBAL UNITS WITH EQUATED DEGREES OF LEARNING[1]

JOHN P. HOUSTON

University of California, Los Angeles

The short-term retention of 5- and 6-word units was tested at 0-, 1½-, 3-, and 6-sec. retention intervals. Degree of learning was equated through manipulation of presentation times. The 5- and 6-word units were presented for 1½ and 2½ sec., respectively. The results indicated that degree of learning was less than 100%, that the 5- and 6-word units did not differ in terms of degree of learning, and that delayed retention of the 5-word units was superior to that of the 6-word units.

One of the products of the recent upsurge in the study of short-term memory (STM) has been the hypothesis that the rate of forgetting of a verbal unit is a function of the number of elements in that unit (Melton, 1963). The bulk of the data supporting this proposal has been obtained with the technique developed by Peterson and Peterson (1959) in which *S*, following a single presentation of the to-be-recalled unit, counts backwards by 3's or 4's until he receives a cue to recall the verbal unit. The consistent finding has been that the slope of the retention curve increases as the number of elements in the recall unit increases. Degree of learning, as indexed by performance on an immediate retention test, has generally approximated 100% for units of all tested sizes (Melton, 1963). While these data have been well received they have not been free from criticism. In particular, there has been some question as to whether the observed differences in retention curves have been due to the effect of number of elements or to unobserved and uncontrolled differences in degree of learning. Keppel and Underwood (1962) have pointed out that differences in degree of learning may be underestimated when immediate retention, as the measure of degree of learning, is essentially 100%. Even though units of differing numbers of elements are recalled perfectly in an immediate retention test the degree of learning of the smaller unit may actually be greater than that of the larger unit and may, therefore, account for the superior recall of the smaller unit at longer retention intervals.

The present experiment was designed to test short-term retention of units of varying numbers of elements when degrees of learning were equated through manipulation of stimulus-presentation times. Five- and six-word units were tested at an immediate and at three delayed retention intervals. Stimulus-presentation times were chosen such that the immediate recall measures of the five- and six-word units were equal and significantly below 100%.

METHOD

Subjects.—The *S*s were 192 University of California undergraduates whose participation was in fulfillment of a course requirement. Most of the *S*s had had some experience with memory-drum learning but none had engaged in a short-term memory study.

Materials.—Two sets of six four-letter single-syllable nouns were developed (FLAG,

---

[1] This investigation was carried out during the tenure of the author's United States Public Health Service postdoctoral fellowship at the Institute of Human Learning, University of California, Berkeley.

HAIR, DESK, BIRD, NOON, CAKE and SNOW, HORN, FACE, LAMP, KING, MEAT). All nouns occurred at least 50 times per million in the Thorndike and Lorge (1944) word count. According to Deese (1962) and the Russell and Jenkins (1954) norms, intraset word-association strength was minimal. The items in each set possessed no apparent meaning similarity. Repetition of letters in the same positions across the nouns within each set was minimal.

Six orders of each set of nouns were developed such that each noun occurred once in each of the six positions. These 12 sequences of six nouns were assigned equally frequently to *S*s in each of the six-word retention conditions. The five-word units were taken from the same pool of nouns by developing, for each set of nouns, six orders of five nouns such that all six nouns occurred once in each of the five positions. These orders were assigned equally frequently to *S*s in each of the five-word retention conditions.

Procedure.—The design called for five- and six-word units to be tested at each of four retention intervals. These eight retention conditions were randomized 24 times such that each condition occurred once in each of 24 blocks. The *S*s were assigned to these blocks in the order of their appearance in the laboratory for a total of 24 *S*s per condition. The *S*s were tested only once.

Pilot data indicated that 2.5-sec. and 1.5-sec. presentation times would result in approximately 75% correct immediate retention of the six- and five-word units, respectively. Accordingly, these presentation times were employed in presenting the five- and six-word sequences to *S*s with a Stowe memory drum. The *S*s were instructed to read the words aloud from left to right when they appeared. Immediately following stimulus presentation the delayed retention groups were shown a 3-digit number for 1 sec. The *S*s read the number aloud and then counted backwards by 3's until three asterisks appeared. The asterisks were *S*'s cue to recall the words in order. The delayed retention intervals were 1.5, 3, and 6 sec. Immediate recall groups were presented the asterisks immediately following stimulus presentation without interpolated counting. All *S*s were allowed 30 sec. for recall.

RESULTS

The pattern of the results obtained with the two sets of nouns did not differ so they were pooled and analyzed together. Figure 1 shows the percentage of the *S*s in each retention condition who displayed total recall (correctly recalled all of the words in the proper order). The immediate or 0-sec. total recall frequencies for the five- and six-word units (18 and 19, respectively) did not differ significantly, χ^2 (1) <1. The exact probability of these immediate total recall frequencies occurring by chance, if degree of learning had been 100%, was less than .01. These data indicate that degree of learning was less than 100% and that the five- and six-word units did not differ in terms of this factor.

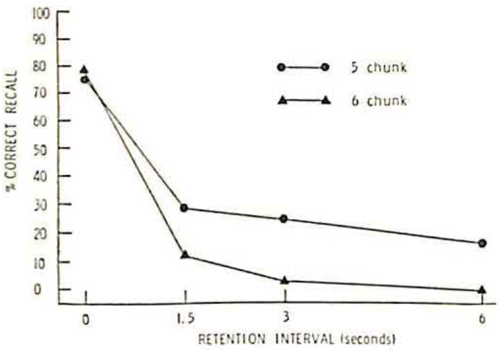


FIG. 1. Percent total recall as a function of unit size and retention interval.

The total recall frequencies for the five- and six-word units were summed over the three delayed-retention intervals. Total recall of the five-word units over the three delayed intervals (17 times) was more frequent than total recall of the six-word units (four times), χ^2 (1) = 8.03, $p < .01$. The relative infrequency of total recall in the delayed retention conditions ruled out the possibility of comparing five- and six-word recall at each interval, but Fig. 1 indicates that the difference between the curves was fairly regular across intervals.

A second less stringent measure of recall was also employed. The percentage of the words recalled correctly, regardless of the order of their recall, was tabulated for each *S*. The percent recall means for the 0-, $1\frac{1}{2}$-, 3-, and 6-sec. intervals were 98.61, 77.75, 72.88, 68.03 and 97.50, 86.66, 78.33, 78.33 for the six- and five-word units, respectively. An analysis of variance applied to the delayed retention data indicated that percent recall of the five-word units was significantly better than percent recall of the six-word units, $F\ (1,\ 138) = 7.87$, $MS_{\text{error}} = 3.10$, $p < .01$. Combined percent recall of the five- and six-word units decreased as the retention interval increased, $F\ (2, 138) = 3.39$, $MS_{\text{error}} = 3.10$, $p < .05$. The interaction was not significant, $F < 1.0$.

Discussion

The principal finding of the study was that short-term retention of five-word units was significantly better than retention of six-word units when degrees of learning had been equated through manipulation of presentation times. This result supports the contention that STM of verbal units is a function of the number of elements in the to-be-recalled units, and indicates that uncontrolled differences in degrees of learning cannot completely account for observed differences in the retention of such units. Intraitem interference (Melton, 1963) appears to be a major factor in the differential short-term forgetting of these five- and six-word units.

The significant difference obtained with the percent recall measure indicates that the total recall difference was not merely due to differences in the associative connections between the words. The availability of the words, apart from sequential dependencies between them, was affected by the numbers of elements in the to-be-recalled units. Thus both "context associations" and "specific associations" seem to be affected by the number of elements in the test units. It should be kept in mind, though, that the percent recall measure may have been subject to uncontrolled differences in degree of learning as immediate percent recall of both the five- and six-word units approximated 100% (98% and 97%, respectively).

The retention curves obtained in the present study are considerably lower than those predicted and obtained by Melton (1963). This discrepancy may well be due to the fact that degree of learning in the present study, as measured by immediate retention, was from 20 to 25% lower than the degrees of learning reported by Melton. It could also reflect the differences in retention materials.

While the present study supports the notion that STM is a function of the number of elements in the recall unit it may also be instructive in that it underscores at least two methodological problems associated with the Peterson STM technique. The first involves the alternative merits of multiple vs. single retention tests. While multiple testing reduces the number of *S*s needed, and allows the use of more powerful statistics, it also necessitates a consideration of the interfering influence of the number of previously learned units upon the retention function. On the other hand, the present study indicates that the use of single tests, while eliminating experimentally established PI effects, is extremely expensive in terms of *S* time. The study also indicates that the control of degree of learning is not a simple matter. The technique employed here has several drawbacks. First, it requires extensive pretesting to determine the appropriate presentation times. In addition, at least with the materials employed, it does not allow the use of recall units of less than five words. Pretesting with less than five nouns indicated that there was no presentation time which, while allowing *S* to say the words aloud, would result in less than perfect immediate recall. With less than five words *S* could either recall the words correctly in an immediate retention test or he could not say them all aloud in the allotted presentation time. Thus, although the notion of equating degrees of learning through manipulation of presentation times appears to be a promising technique, it is not without its limitations.

REFERENCES

DEESE, J. Word-association norms. Unpublished manuscript, Johns Hopkins University, 1962. (Mimeo)

KEPPEL, G., & UNDERWOOD, B. J. Proactive inhibition in short-term retention of single items. *J. verbal Learn. verbal Behav.*, 1962, **1**, 153–161.

MELTON, A. W. Implications of short-term memory for a general theory of memory. *J. verbal Learn. verbal Behav.*, 1963, **2**, 1–21.

PETERSON, L. R., & PETERSON, M. J. Short-term retention of individual verbal items. *J. exp. Psychol.*, 1959, **58**, 193–198.

RUSSELL, W. A., & JENKINS, J. J. Language in behavior: The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Word Association Test. Technical Report No. 11, 1954, University of Minnesota, Nonr 66216, Office of Naval Research.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, Bureau of Publications, 1944.

(Received April 10, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 1, 79–82

SCALING MUSICAL PREFERENCES[1]

SOON DUK KOH

Hastings State Hospital, Minnesota

2 psychophysical scaling procedures, category judgment and magnitude estimation, were applied to affective judgments of a total of 111 musical excerpts. The affective category scale was found to be a roughly logarithmic function of the affective magnitude estimation scale. To this extent the affective scale behaves as an "intensity" or prothetic continuum. The category-magnitude relation is invariant under differences in S's education, age, sex, and alcoholic deterioration.

In psychophysics, Stevens' (1957) power law relating sensations to stimuli and the nonlinearity between the category and magnitude scales (Stevens & Galanter, 1957) are observed only on continua that involve intensity or quantity (prothetic continua). Most psychophysical continua belong to this class. On continua having to do with kind or quality (metathetic continua) the power law may or may not hold, but the relation between category and magnitude scales is linear. Affective judgments of music or paintings are often regarded as qualitative judgments in a semantic sense, but the intensity of the affective judgments on a given continuum may be prothetic. The present study applies the measurement paradigm of psychophysics to preference judgments of musical selections in order to reveal the nature of the continuum.

The measurement of perceptual continua by psychophysical scaling procedures has become a lively topic since Stevens' massive attack on the traditional Fechner-Thurstone approaches (Stevens, 1957, 1961, 1962; Stevens & Galanter, 1957). Magnitude estimation is one of the newly developed "ratio" scaling methods in which *S* is instructed to assign a number to each stimulus presentation so that the numbers are proportional to the subjective intensity experienced by *S*. Category scaling, on the other hand, is an older method; in this method *S* is, for instance, instructed to assign the lowest category to the weakest stimulus, the highest to the strongest stimulus, and to assign the intermediate categories in such a way that his subjective impression of the distances between successive categories is the same. The category scale is, therefore, at best an interval scale. When the two kinds of scales employing the same stimuli are compared with each other the category scale usually takes a form that is intermediate between a logarithmic function and the form of the corresponding magnitude scale, provided the continuum is prothetic. This means that stimuli whose differences are reported to be subjectively equal on the category scale are separated by progressively larger distances on the magnitude scale. Both types of scales have reasonably high repeatability, and yet their relationship is always nonlinear. The reason for this nonlinearity has caused much speculation, and the problem may remain unsolved until an independent measure of sensation or a new idea arises.

The purpose of this study is

---

[1] This investigation was supported by Public Health Service Research Grant M-05306 from the National Institute of Mental Health. It was conducted at the State Hospital, Jamestown, North Dakota.

to examine the possible extension of Stevens' magnitude-estimation method to complex affective judgments and to determine the relation between the category and magnitude-estimation scales for affective judgment of musical excerpts. If Stevens' ratio methods, which are comparatively new, are applicable to affective judgments it would contribute significantly to experimental esthetics. If affective judgments are like psychophysical judgments and if affective category and magnitude scales are shown to be nonlinearly related to each other, then affective judgments are presumably prothetic rather than metathetic. This outcome should affect the choice of scaling method and the interpretation of affective data in future scaling procedures.

METHOD

Two groups of musical excerpts were used: 51 vocal and 60 piano pieces. The vocal excerpts were selected from records and included opera singing, folk songs, college songs, and stage music. Each excerpt lasted about 60 sec. The piano excerpts were all from classical music, and were selected and recorded by a pianist. Each excerpt lasted about 15 sec. Both vocal and piano excerpts were separately presented to *S*s by a tape recorder, in groups of about 30–60 *S*s. The interval between the individual excerpts was about 10 sec. Because of the instrumental backgrounds, differences in solo, duet, chorus, etc. the vocal selections provided a wider variety than the piano selections.

For the category scale *S* was instructed to assign one of the nine ordered response categories, ranging from most pleasant, extremely, moderately, mildly, indifferent, to most unpleasant, to each excerpt on the basis of his affective impression of pleasantness-unpleasantness. The category assignment for each excerpt by *S* was transformed into the normalized distribution, and the median category assignment for each excerpt was calculated. Since the scale values found by this procedure were very similar to those found by simply calculating the mean category number, the latter procedure was adopted.

For the method of magnitude estimation, a standard excerpt was played at the beginning of the experiment and written instructions were given as follows.

> Let us call your *preference level* or your *pleasantness level* for the standard musical selection 10. Your task is to tell your preference level for other musical selections, proportional to this standard. In other words, the question is: If the standard music is called 10, what would you call the other musical selections: For example, if another musical selection is felt to be 7 times more pleasant than the standard, write down 70. If it is felt to be one fifth as pleasant, jot down 2; if a twentieth as pleasant, jot down 0.5; if it is equal to the standard, jot down 10, etc. Try not to worry about being consistent; try to give the appropriate number to each musical selection regardless of what you have called some previous stimulus.

The geometric mean of the proportional numbers given by *S* was calculated for each excerpt. Both methods, category and magnitude estimation, were administered in a single sitting, presenting the same 51 vocal or 60 piano excerpts twice. The *S*s for the vocal excerpts were 56 college men, 52 college women, and 51 men hospitalized for alcoholism. The *S*s for the piano excerpts were 55 college men, 51 college women, and 65 male alcoholic patients. The students were mostly freshmen and sophomores drawn from four different colleges and the patients were from a state hospital. The average age for the patients was 42 yr. and their education 10 yr. About 90% of the patients' occupations before the hospitalization were semiskilled, farmer, and laborer.

RESULTS AND DISCUSSION

The results of the six experimental groups are summarized in Fig. 1. The data were plotted on semilog coordinates, and the line of best fit was computed by the method of least squares. According to this line, the category value C is related to the magnitude-estimation value M by the following equations:

With the vocal excerpts, college men, $C = 3.39 + 2.93 \log M$; college women, $C = 3.89 + 2.88 \log M$; patients, $C = 4.08 + 2.35 \log M$.

With the piano excerpts, college men, $C = 2.13 + 3.31 \log M$; college women, $C = 2.35 + 3.46 \log M$; patients, $C = 1.16 + 3.98 \log M$.

The relations between the category and magnitude scales are all approximately linear on the semilog coordinates; i.e., the affective category scale is a function of the logarithm of the affective magnitude-estimation scale. The form of relation between the two scales is strikingly invariant in spite of the marked differences of age, education, occupation, and even the alcoholic pathology of *S*. It is also invariant under the difference of vocal vs. piano excerpts. The slopes for the piano excerpts are consistently steeper than those for the vocal excerpts, but are approximately equal for a given type of excerpt. The product-moment correlations between the category scale and the logarithm of magnitude-estimation scale range from .90 to .96.

It is noted above that the category scale is usually not logarithmic. As a matter of fact, some of the data in Fig. 1 show evidence of a slight upward concavity. This curvature is slight, but such a curvature is always

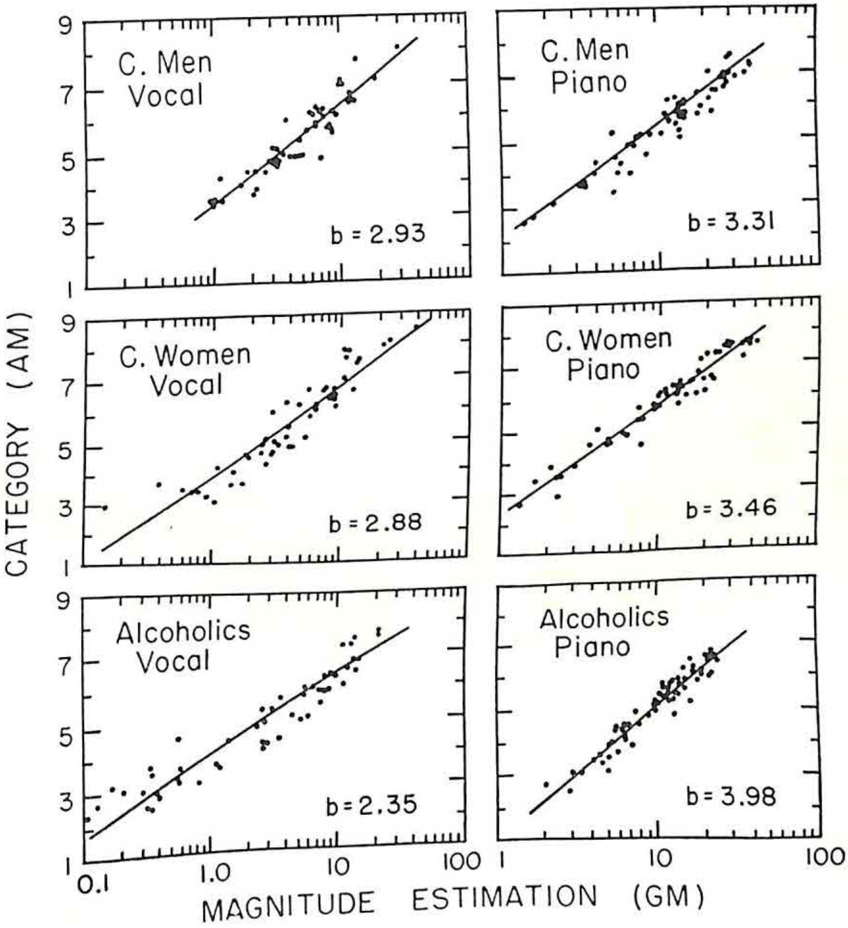


FIG. 1. Category judgments and magnitude estimations of affective impressions for 51 vocal and 60 piano excerpts. (The arithmetic means of the category judgments are plotted against a logarithmic scale derived from the geometric means of the magnitude estimations. The *b* values refer to the slopes of the lines fitted to the data by least squares.)

present in psychophysical data when the "pure" form of the category scale is plotted against the logarithm of the magnitude-estimation scale (Stevens & Galanter, 1957). It seems that this curvature becomes more evident in the present data when the affective range is sufficiently wide, as shown by the judgments of the vocal selections by the women and the alcoholics. When the range is short and the variability fairly high, the upward curvature may be less evident. Stevens and Guirao (1963) show how variability affects the category scale.

The present finding suggests that affective judgment for music exhibits a feature found on prothetic continua in psychophysics. Ekman and Kuennapas (1962a, 1962b) measured esthetic values of handwritings and drawings and found a logarithmic relation between the scales derived from the Thurstonian method of pair comparison and the method of ratio estimation. With drawings they also obtained a category scale that was similar to those obtained here.

Stevens (1957) has proposed four criteria that are relevant to the distinction between the prothetic and metathetic classes of perceptual continua, viz., the subjective size of the jnd, the form of the category scale, the time error, and hysteresis. The nonlinear relation of the two scales found in this study satisfies one of the four criteria for prothetic continua. In a series of studies Koh (1962, 1963) found clear evidence of a time error in affective judgments of musical excerpts. The prothetic nature of affective judgments seems, therefore, to be still more plausible.

If the determinants of perception are classified into stimulus parameters and personality parameters, or "autochthonous" and "behavioral" determinants, the affective judgments of musical excerpts may involve more interaction of these two determinants and more behavioral variables than psychophysical judgments. In view of the immediate payoffs involved in hearing the musical excerpts, the possible idiosyncrasy of musical preference, and the more-or-less ambiguous stimulus features, it is interesting to find that stimulus-bound, autochthonous features were still present in the affective judgments. The result also indicates the usefulness of the method of magnitude estimation for the study of complex processes like affective judgments. Since Stevens' scaling methods have been explored mostly with college students the applicability of this method to the mental patient and to the low educational group is also encouraging.

REFERENCES

EKMAN, E., & KUENNAPAS, T. Measurement of aesthetic value by "direct" and "indirect" methods. *Scand. J. Psychol.*, 1962, **3,** 33–39. (a)

EKMAN, E., & KUENNAPAS, T. Scales of aesthetic value. *Percept. mot. Skills*, 1962, **14,** 19–26. (b)

KOH, S. D. The Fechnerian time error and Pratt's time error in affective judgments of musical excerpts. *Amer. Psychologist*, 1962, **17,** 369. (Abstract)

KOH, S. D. The effect of interpolated stimuli upon affective comparisons of musical excerpts. *Amer. Psychologist*, 1963, **18,** 437. (Abstract)

STEVENS, S. S. On the psychophysical law. *Psychol. Rev.*, 1957, **64,** 153–181.

STEVENS, S. S. To honor Fechner and repeal his law. *Science*, 1961, **133,** 80–86.

STEVENS, S. S. The surprising simplicity of sensory metrics. *Amer. Psychologist*, 1962, **17,** 29–39.

STEVENS, S. S., & GALANTER, E. H. Ratio scales and category scales for a dozen perceptual continua. *J. exp. Psychol.*, 1957, **54,** 377–411.

STEVENS, S. S., & GUIRAO, M. Subjective scaling of length and area and the matching of length to loudness and brightness. *J. exp. Psychol.*, 1963, **66,** 177–186.

(Received April 10, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 1, 83-86

VIGILANCE PERFORMANCE WITH A QUALITATIVE SHIFT IN REINFORCERS[1]

WILLIAM BEVAN AND EDWARD D. TURNER

Kansas State University

This experiment examined the effect of a qualitative shift in reinforcers upon performance in a simple vigilance task. Ss in 5 independent groups reported the absence of a tone as it occurred from time to time in a series presented at regular intervals for approximately 1 hr. An unreinforced group performed at about the 50% detection level throughout. 2 groups, 1 receiving a penny for each correct response throughout and 1 receiving a mild electric shock for each error, performed equally well and significantly better than the unreinforced group. The best performance occurred in 2 groups experiencing a qualitative shift in reinforcers from the 1st to the 2nd halves of the experimental task. Furthermore, the direction of change—from coins for correct responses to shock for errors or vice versa—had no demonstrable relationship to the facilitation of performance. Since shock was involved and it was necessary to allow S prior knowledge of his reinforcement treatment, anticipation of the reinforcement shift became an inevitable adjunctive condition. The data indicate that such anticipation was effective in facilitating performance both before and after the actual occurrence of the shift.

Recent experimental studies (Adamson, Bevan, & Maier, 1961; Bevan & Adamson, 1960; Black, Adamson, & Bevan, 1961) have dealt with the effectiveness of a particular reinforcer as a function of the intensive level of prior stimulation and have given rise to an interpretation of reinforcement phenomena in terms of adaptation-level theory (Bevan, 1963; Bevan & Adamson, 1962). Reinforcement contrast is a phenomenon readily accommodated by this conceptual approach. The present experiment attempts to extend its rationale and to explore the effect of a nonrandom shift in the *qualitative* nature of the reinforcer. Here two possibilities seem reasonable: (*a*) Qualitative differences in reinforcer are incidental to intensive differences and thus qualitative shifts will influence performance efficiency only when an intensive shift is also involved. (*b*) Extending the "law of affective contrast" (Bacon, Rood, & Washburn, 1914), the affective properties of stimuli are important to their role as reinforcers and thus a qualitative shift should eventuate in a contrast effect even in those cases where no intensive differences are involved.

METHOD

Design.—The experimental task involved a series of short tones at regular intervals with occasional missing signals and instructions to report the latter by pressing a telegraph key. Reinforcement consisted of mild shock for errors and/or pennies for correct detections. Because shock was specified for certain of the *S*s, each *S* scheduled for reinforcement received at the outset of testing information about reinforcement appropriate to his experimental assignment. He was told that he would receive shock for errors throughout, pennies for correct detections throughout, or one form with a shift to the other sometime during testing.

The design involved five independent groups:

---

[1] This experiment was performed under Contract Nonr-3634(01) between Kansas State University and the Physiological Psychology Branch, Office of Naval Research.

Group I. The *S*s in the control group received no reinforcement throughout the test sequence.

Groups II_r and II_p. The *S*s in these two groups received one type of reinforcement, reward or punishment, throughout the test sequence. They represent the traditional reinforcement condition.

Groups $\text{III}_{r,p}$ and $\text{III}_{p,r}$. The *S*s in these two groups received one type of reinforcement, reward or punishment, during the first half of the test session, and the other, punishment or reward, during the second half.

The present vigilance task was chosen as the experimental vehicle after preliminary work with a variety of materials because it was sufficiently difficult to allow for improvement in performance with reinforcement and at the same time yielded stable enough performance to allow for a competent assessment of the effect of the qualitative shift.

Subjects.—The *S*s were 50 introductory psychology students divided at random into five groups of 10 each, with an approximately equal number of men and women in each group.

Apparatus.—The auditory signal was a 1,000-cps tone, $\frac{1}{2}$ sec. long, with a rise and decay time of $\frac{1}{4}$ sec. The interstimulus interval was 1 sec. There were 32 missing signals scheduled randomly throughout the test sequence at $\frac{1}{2}$, 1-, 2-, and 4-min. intervals. This program, on high-fidelity magnetic tape, was reproduced on a Webcor recorder (Model No. EP 2101-1) and played through Willson sound-barrier earphones. The *S*'s detection responses, along with the occurrence of the signals and the presentation of the reinforcement, were recorded on an Esterline-Angus recorder. Pennies for correct responses were delivered from a dispenser designed and constructed at the laboratories of the Menninger Foundation. The 0.2-sec. 3-mamp. shock for errors was generated by an Applegate electronic stimulator (Model 226B).

Procedure.—The *S*s were tested individually in a dimly lit, sound-insulated room. First, a series of pretest signals was used to establish *S*'s 50% detection threshold. The *E* then explained the task and the particular condition of reinforcement to be applied. Next a 2-min. series of tones, all clearly above threshold and with the third, fourth, fifth, fourth, and third tones missing, was presented to further familiarize *S* with the nature of the task and *S*'s questions were answered. Finally, the experimental program was run at the 50% detection intensity. The duration of the experimental session was approximately 1 hr.

RESULTS AND DISCUSSION

Performance on the auditory vigilance task.—Curve I in Fig. 1 presents the average number of errors made by the unreinforced group during the first and second halves of the test task. Its most striking characteristic is that, unlike functions commonly obtained with simple monitoring tasks, it shows no performance decrement with the passage of time. The average number of detection errors was 7.9 for the first 16 missing signals and 7.3 for the last 16. Thus, performance was approximately at the 50% detection threshold throughout. Analysis of the program into successive blocks of 4 missing signals each showed no systematic shifts in performance from block to block, $F_{\text{between blocks}}$ (1, 72) = .26, $p > .05$.

Such results appear to be consistent with Deese's (1955) "activation hypothesis" which views the problem of maintaining vigilance as one of monitoring a background of sensory input.

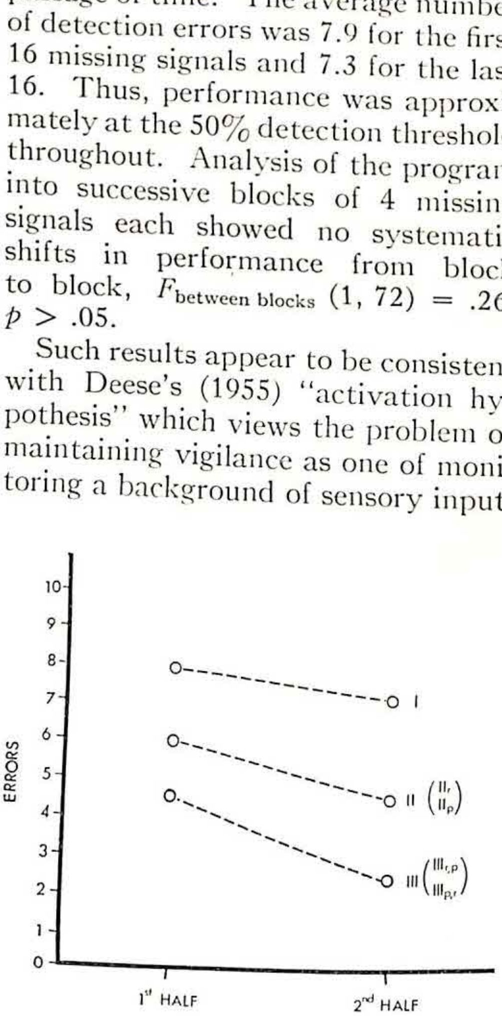


FIG. 1. Average errors on a simple vigilance task under different conditions of reinforcement. (I is the control group. II represents combined data for groups receiving either reward or punishment throughout the task. III represents combined data for groups shifting from reward to punishment or punishment to reward on the second half of the task.)

The present task, the complement of the typical vigilance program with its infrequently occurring signals, would appear to meet this requirement.

Effect of reinforcement upon vigilance performance.—Curves II and III indicate clearly that reinforcement facilitated vigilance performance, $F_{\text{I vs. II, III}}$ (1, 45) = 11.52, $p < .001$. In this respect, the present results join a growing literature. Under the particular conditions of simple reinforcement of this experiment (Cond. II), reward and punishment—i.e., pennies and mild shock—were equally facilitative, $F_{\text{IIr vs. IIp}}$ (1, 45) < 1.0, $p > .05$. This represents a fortunate circumstance, for it permits *E* to test for a possible contrast effect, unconfounded by intensive differences between the reinforcers involved in the qualitative change.

Performance under conditions of simple and changed reinforcement.—A comparison of Group II and Group III performance on the *second* half of the task indicates the superiority of the latter and suggests a contrast-like effect. It would appear that a change in affective quality associated with reinforcement leads to enhanced affective tone and better performance. However, inspection of Curve III forces rejection of this simple conclusion, for the groups subjected to the reinforcement shift were also superior in the *first* half of the task. Errors averaged significantly fewer throughout than in II, $F_{\text{II vs. III}}$ (1, 45) = 4.45, $p < .05$, and there was no difference in the slopes of the two reinforcement curves across halves, F_{slope} (1, 45) = 2.82, $p > .05$. Furthermore, the direction of qualitative shift made no reliable difference in its effect upon performance, $F_{\text{III}_{\text{r,p}}\text{ vs. III}_{\text{p,r}}}$ (1, 45) < 1.0, $p > .05$.

These data prompt the conclusion that an additional ingredient had been added to the reinforcement situation from the outset. Review of procedure readily identifies this ingredient: anticipation of shift. Because electric shock was included as one form of reinforcer, it was unavoidable that *S*s scheduled to receive shock be informed of this condition. The requirements of good experimental procedure then demanded that all groups scheduled for reinforcement be given prior information about their treatment. Under these circumstances, it seems reasonable to interpret the superior performance of Group III to result from a summation of physical reinforcement and anticipation. By the same token, there is an inevitable confounding between the anticipation of and the actual occurrence of a shift that precludes any statement about the influence of the shift per se. At the same time, the marked superiority of the groups experiencing shift over the unreinforced group is an empirical datum worthy of note.

Two conditions militate against the appearance of a qualitative contrast effect: the low initial level of errors in the III groups and the averaging of performance across each half. In order to sharpen up the preshift-postshift difference, improvement in performance from the last block of four test trials before the shift to the first four trials after the shift was compared for Cond. II and III. This comparison failed to reveal any qualitative contrast effect, $F_{\text{diff.}}$ (1, 45) = 2.00, $p > .05$.

Since anticipation of reinforcement shift confounds the assessment of the shift itself, an experimental situation allowing for the introduction of a change in reinforcers without *S*'s prior knowledge provides a better test of the contrast effect. This is most readily accomplished with animals and such a study is now underway in our laboratory.

REFERENCES

Adamson, R., Bevan, W., & Maier, B. Shifts in bar pressing as a function of alternating extinction and reinforcement. *J. gen. Psychol.*, 1961, **64**, 147–152.

Bacon, M. M., Rood, E. A., & Washburn, M. F. A study of affective contrast. *Amer. J. Psychol.*, 1914, **25**, 290–293.

Bevan, W. The pooling mechanism and the phenomena of reinforcement. In O. J. Harvey (Ed.), *Motivation and social interaction*. New York: Ronald Press, 1963. Pp. 18–34.

Bevan, W., & Adamson, R. Reinforcers and reinforcement: Their relation to maze performance. *J. exp. Psychol.*, 1960, **59**, 226–232.

Bevan, W., & Adamson, R. Internal referents and the concept of reinforcement. In N. F. Washburne (Ed.), *Decisions, value and groups*. Vol. 2. New York: Pergamon Press, 1962. Pp. 453–472.

Black, R., Adamson, R., & Bevan, W. Runway behavior as a function of apparent intensity of shock. *J. comp. physiol. Psychol.*, 1961, **54**, 270–274.

Deese, J. Some problems in the theory of vigilance. *Psychol. Rev.*, 1955, **62**, 359–368.

(Received April 11, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 1, 87–93

EFFECT OF SUCCESSIVE ADDITION OF STIMULUS ELEMENTS ON PAIRED-ASSOCIATE LEARNING[1]

SAM C. BROWN
Kansas State University

WILLIAM F. BATTIG
University of Maryland

AND RICHARD PEARLSTEIN
University of Virginia

Successive addition of nonoverlapping stimulus letters, following attainment of low or high learning criteria for each pair, significantly facilitated paired-associate learning, both with interpair variation (mixed) and consistency (unmixed) in sequence of letter addition, as compared with the typical (control) condition wherein all 3 stimulus letters were presented from the outset. Learning criterion had no overall effect on performance, but interacted with sequence of unmixed letter addition. Postexperimental recall of stimulus letters showed lesser but still substantial recall of added letters by successive-addition groups, especially under high-criterion conditions. The results are interpreted in terms of a reduction in interference, and more effective use of added letters to overcome such interference at intermediate stages of learning, under successive-addition conditions.

Previous research concerned with the role of serial-position cues in paired-associate learning (Battig, Brown, & Nelson, 1963; Brown & Battig, 1962) has indicated these cues to develop functional stimulus properties, in addition to rather than in place of the presented stimulus term, primarily during and following intermediate stages of learning (e.g., after a given pair has been responded to correctly). On the basis of Brown's (1964) demonstration that interpair interference develops maximally at intermediate stages of paired-associate learning, it would appear that the belated facilitative effect of these added serial-position cues may have been to help overcome or reduce such interpair interference. More general indications that additional stimulus cues become learned as a means of combating interpair interference have come from (*a*) Johnson's (1964) demonstration that amount of stimulus learning of nonsense shapes, as indexed by subsequent recognition accuracy, becomes markedly greater with increasing paired-associate task difficulty and interpair interference, and (*b*) the incidental finding of both Sundland and Wickens (1962) and Newman (1963) that an additional stimulus cue, although leading to poorer initial performance than for a single-cue group, produced slightly better performance by the final test trial.

The foregoing analysis deviates somewhat from Underwood's (1963) recent discussion of stimulus selection, which implies that instead of associating the response with an entire multielement stimulus, *S*s will typically select the one of these elements that provides the most effective functional stimulus. According to this view, the addition of new elements to an already adequate one-element stimulus, at or after intermediate stages of learning, should produce little or no learning or use of these added elements. Based on the present analysis, however, it would follow that the introduction of new elements at intermediate stages of learning, where interpair interference is maximal, should produce more effective learning and use of these added elements to combat interpair interference, than if the added elements are present initially so as to represent an additional source of interference per se (Newman, 1963). To evaluate this possibility, the present study compared performance using constant three-letter stimulus terms, against conditions beginning with a single-letter stimulus and successively adding second and third letters upon the attainment of specified learning criteria.

---

[1] This investigation was conducted at the University of Virginia, supported by a Public Health Service predoctoral fellowship to the first author, and by Public Health Service Research Grants M-05769-02 and HD-00929-03 to the second author. Analysis of the data and preparation of the manuscript were completed during the second author's tenure as a Visiting Scholar at the Institute of Human Learning, University of California, Berkeley.

METHOD

Subjects and materials.—Each of 120 paid *S*s, all summer session students or affiliates at the University of Virginia, learned the same six-pair list (MJX-BAG; PVN-BIG; SZL-BIN; CFK-TAG; RBW-TAN; and HQY-TIN). All stimulus terms were consonant syllables of zero frequency of occurrence in the English language and minimal interletter association value (Underwood & Schulz, 1960), with no letter being duplicated within the list. The six response terms, maximally alike in letter composition, were all highly meaningful words of 99–100% Archer (1960) association value, and with Thorndike-Lorge (1944) frequencies of at least six per million.

Design.—According to a predetermined unsystematic sequence based on order of appearance in the laboratory, *S*s were assigned to five basic groups of 24 each, which differed only with respect to the particular stimulus letters presented at various stages of learning. For the *control* group, the typical recall paired-associate procedure was employed, with all three stimulus letters presented on each trial. In the other four groups, however, only one stimulus letter was presented for each pair on initial trials, with the other two letters being successively added on subsequent trials. For the *mixed* group, one of the six pairs represented each of the six possible sequences of addition of the letters appearing in the left (L), middle (M), and right (R) positions in the syllable (LMR, LRM, MLR, MRL, RLM, and RML). Through a Latin-square counterbalancing arrangement, each of the six sequences was used equally often within each pair. For each *S* in the remaining three *unmixed* groups, all six pairs involved the identical sequence of letter addition, these being LMR, MRL, and RLM, respectively, for the three groups. Within each mixed and unmixed group, half of the *S*s constituted a *low-criterion* condition, wherein the second stimulus letter was added to each pair on the trial immediately following the first correct response to the one-letter stimulus, and the third letter was introduced after the trial of the first correct response to the two-letter stimulus. For the *high-criterion* *S*s, addition of the second and third letters to each pair was correspondingly contingent upon attainment of a criterion of three consecutive correct responses to that pair. Learning for all *S*s, including the control group, was continued to the same criterion of three successive errorless trials on the entire list with the three-letter stimulus.

Each group and subgroup was equally subdivided according to whether *S*s spelled or pronounced their responses, and half of the *S*s under each condition were run by each of two *E*s. Since neither of these variables significantly affected performance, either alone or in interaction, they have been ignored in presentation of the results.

Procedure.—Each recall-method trial began with the manual presentation of the six pairs typed on 3 × 5 in. cards at a 3-sec. rate, followed after 15 sec. by the presentation in a different serial order of the six stimulus terms alone at a 3-sec. rate for attempted response-term recall. The intertrial interval was 30 sec. Following the final criterion trial, each *S* was given a sheet containing the six response terms, and attempted to write down on a line to the left of each response term as many of the three correct stimulus letters as possible within a 2-min. time limit. Subsequent relearning trials were then given to a criterion of one errorless trial, in which only the initial one-letter stimulus was presented to the four successive-addition groups, and the first (L) letter to the control group. Each *S* was then questioned as to which of the three stimulus letters he had used for the various pairs.

RESULTS

As shown in Table 1, the successive-addition procedure necessarily required significantly more trials than

TABLE 1

MEAN TOTAL TRIALS AND ERRORS TO CRITERION FOR GROUPS AND SUBGROUPS

	Group					
	Control	Mixed	RLM	LMR	MRL	Total
Trials						
Low		12.8	9.2	10.5	13.3	11.4
High		15.8	17.6	16.3	13.5	15.8
Total	10.5	14.3	13.4	13.4	13.4	
Errors						
Low		17.8	9.9	16.3	26.7	17.7
High		18.6	20.8	17.6	10.8	16.9
Total	23.9	18.2	15.3	17.0	18.7	

for the control group, $F\ (1, 102) = 7.97$, $p < .01$, but also produced consistently fewer total errors. Although the error difference in favor of the combined successive-addition over control procedures fell short of significance, $F\ (1, 102) = 3.50$, $p < .10$, two of the individual unmixed groups (LMR and RLM) were significantly ($p < .05$) below the control group in total errors. A further breakdown into *before errors* (preceding the first correct response to each pair) and *after errors* (errors after each pair had been responded to correctly for the first time) indicated the former measure to be primarily responsible for the difference between successive addition (9.6) and control (15.6) conditions, $F\ (1, 102) = 12.95$, $p < .001$. However, insignificantly ($F < 1$) fewer after errors were also made by successive-addition (7.6) than control groups (8.3), even though for the former groups this measure included those errors produced by the delayed introduction of the second and third stimulus letters.

Similarly, insignificantly ($F < 1$) fewer total errors were made by the combined high- than low-criterion subgroups, although significantly more trials were necessitated under high- than low-criterion conditions, $F\ (1, 102) = 19.83$, $p < .001$. Low-criterion trials, however, were not significantly above the control group ($F < 1$). As shown in Table 1, the high-criterion superiority in total errors was limited primarily to the MRL group, the other three successive-addition groups showing a general slight low-criterion superiority, which led to significant Criterion × Successive-Addition Groups interactions both for errors and trials measures, $Fs\ (3, 102) = 3.12$ and 3.18, $p\text{'s} < .05$. After errors produced a similar interaction ($F = 3.97$, $p = .01$) which reflected not only the sharp MRL high- (3.3) over low-criterion (14.8) superiority, but also a contrasting decrement in high- (11.1) as compared with low-criterion performance (3.8) within the RLM group. The nonsignificance of this interaction for before errors ($F = 1.35$, $p > .25$) indicates the effect to be a genuine result of the experimental manipulations, rather than a sampling error reflecting subgroup differences in learning ability.

As contrasted with the three unmixed successive-addition groups, wherein RLM performance was best and MRL worst, the mixed group made significantly fewer mean errors per pair on the two pairs (LMR and LRM) for which the one-letter stimulus was also the first letter (2.3) than on the other four pairs (3.4), $F(1, 102) = 8.70$, $p < .005$. However, mixed-unmixed differences did not approach significance ($p > .25$) either for the entire list or for pairs constituting the three unmixed sequences, and no other differences between mixed-list sequences were observed ($F < 1$). Moreover, there was no evidence of any Sequence × Criterion interaction ($F < 1$) comparable to that observed between the three unmixed sequences, the corresponding mixed differences being numerically in the opposite direction.

Stimulus recall.—Presented in Fig. 1 for each of the three stimulus letters is the mean percentage correctly recalled by each group. The group differences displayed in Fig. 1 reflect (*a*) the near perfect recall by each of the three unmixed groups of the initial one-letter stimulus (95.5%), with a marked decrement for the added second (44.1%) and third (37.0%) letters, especially for the LMR group, (*b*) a lesser but consistent overall superiority of recall for the L (70.1%) over the M (58.5%) and R (56.5%) letters, and (*c*) more total letters recalled by the control (73.1%) and mixed (66.2%) groups, and less by the LMR group (50.5%), as compared with the MRL (57.2%) and RLM (61.6%) groups. Analyses of variance corroborated each of these effects, as indicated by significant differences between (*a*) first, second, and third added letters, $F(2, 160) = 191.41$, $p < .001$, shown by trend analysis to reflect both significant linear and quadratic components (p's $< .001$) and thus primarily the superiority of first over second and third letters, (*b*) L, M, and R positions, $F(2, 160) = 13.94$, $p < .001$, also with both linear ($p < .001$) and quadratic ($p < .05$) components due to the superiority of L over M and R positions, and (*c*) the five groups, $F(4, 102) = 4.31$, $p < .005$, shown by subsequent analysis to be attributable to the superiority of control and mixed ($p < .025$) and near significant inferiority of LMR ($p < .10$) to the other two unmixed groups. Recall of the initially presented letter was virtually identical for mixed (95.1%) and the combined unmixed groups (95.6%), so that the overall superiority of the former reflected better mixed-group recall of the second (54.9%) and third (48.6%) added letters, as compared with second- (40.5%) and third-letter recall (33.1%) by the unmixed groups. However, neither this interaction $F(2, 160) = 2.79$, $p < .10$, nor its linear component, $F(1, 80) = 3.55$, $p < .10$, achieved statistical significance.

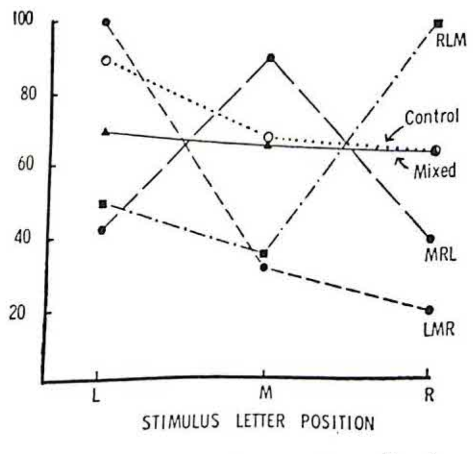


FIG. 1. Mean percent correct recalls of each stimulus letter for each group.

The only other significant stimulus-recall effect was an overall superiority of high- (63.7%) over low-criterion (54.1%) subgroups, $F(1, 102) = 5.34$, $p < .025$. The high-criterion subgroups, although still inferior to the control group, did not differ significantly therefrom, $F(1, 102) = 3.47$, $p < .10$. While the Criterion × Positions interaction failed to approach significance ($F < 1$), and thus is not shown in Fig. 1, the superiority of the

TABLE 2

NUMBER OF *S*S IN EACH GROUP REPORTING USE OF VARIOUS STIMULUS COMPONENTS

Reported Stimulus	Group				
	Control	Mixed	LMR	MRL	RLM
First letter only	3	10	18	14	15
First letter predominant	6	11	4	4	6
Two letters	2	0	1	1	0
All three letters	10	1	0	3	3
Miscellaneous	3	2	1	2	0

high- over the low-criterion subgroups was actually numerically greater for the third added letter (12.2%) than either the first (6.2%) or second (10.4%) letters, even though the learning criterion following the addition of the third letter was identical for the two subgroups.

Relearning.—Relearning performance, with only the initial or single-letter stimulus being presented, was nearly perfect for all groups. All of the 48 high-criterion *S*s performed errorlessly on the first relearning trial, as did 38 of the 48 low-criterion *S*s and 19 of the 24 control *S*s. Analysis of variance of the relearning error scores for the control and low-criterion *S*s revealed only a significant inferiority of the MRL subgroup (1.3) to the other groups combined (0.3), $F\ (1, 62) = 7.29,\ p < .01$.

Postexperimental questioning.—The number of *S*s in each group indicating the use of each number of stimulus letters is presented in Table 2, which suggests a greater reported dependence on a single letter by the successive-addition than by the control *S*s, and also a nearly equal division within the control group between *S*s who consistently selected one or two of the stimulus letters and those who reported making use of all three letters.

DISCUSSION

The present results clearly demonstrate that the presentation throughout learning of all letters of trigram stimuli is less efficient than the present successive-addition procedure, in terms of number of errors required to reach a list-based performance criterion. Since the control-group inferiority was primarily in before errors, it can be attributed largely to the greater initial interference produced by multi- than single-element stimuli (Newman, 1963), probably reflecting *S*s' efforts to learn each of the stimulus letters per se, or to select from among them, rather than learning associations with the response. However, numerically fewer after errors were also made by successive-addition groups, even though the addition of new previously unseen stimulus letters would be expected to represent a major source of interference in itself, and the successive-addition *S*s were in effect required to achieve three learning criteria as contrasted with the single criterion for the control group. Nonetheless, only for the unmixed MRL group was there any indication of significant interference from the successive-addition procedure, and even here it was limited to the low-criterion subgroup. This interaction, such that the MRL high-criterion subgroup showed a marked reduction in errors relative to low-criterion *S*s, suggests that if the initial one-letter stimulus is to appear in the middle of the final three-letter configuration, it must be learned to a high degree in order to avoid interference from the successive addition of both preceding and following letters.

There are several indications that the facilitation for successive-addition *S*s was at least partially due to their more effective use of the added stimulus letters to overcome interpair interference produced from other sources. That substantial interference was involved with the present materials is demonstrated by the fact that of the total errors combined over all groups, 41.7% occurred on pairs after they had been responded to correctly, and also by the recent results of Chang and Shepard (1964). Not only was there substantial learning of the second and third letters under successive-addition conditions, despite their probable lesser availability, but this was significantly greater for mixed groups, for whom the inconsistency in sequence of letter addition from pair to pair would be expected to provide a source of interference not present for unmixed groups. Therefore, the equivalence of mixed and unmixed paired-associate performance implies that the increased mixed-group learning of added stimulus letters had effectively combated the increased interpair interference therein.

Convincing support is thus provided for the generality of the thesis that under conditions of appreciable interpair interference, additional stimulus cues are learned during and following intermediate stages of learning as a means of combating such interference. During the postexperimental questioning, several *S*s in the successive-addition groups indicated their use of the added letters for this purpose. However, substantial recall of added elements was also demonstrated by *S*s who reported trying to ignore them, several of whom expressed puzzlement as to how and why these letters could be successfully recalled.

The group differences in stimulus recall displayed in Fig. 1 are clearly indicative of marked discrepancies between the nominal and functional stimulus in paired-associate learning, in support of Underwood's (1963) analysis. They indicate, however, that Underwood's views concerning stimulus selection may have at best limited generality, since the present *S*s demonstrated substantial learning of additional stimulus elements even under conditions expressly designed to maximize selection of a single-element stimulus. The substantial level of interpair interference under the present conditions is likely responsible for this discrepancy, as well as for the greater dependence upon a single-letter stimulus reported by Maddocks (Underwood & Schulz, 1960) and the greater effects of spatial position of the letter found by Jenkins (1963), inasmuch as both studies probably involved considerably less interpair interference than the present one.

Since none of the letters appeared more than once in the present list of stimulus terms, meaning that each added letter was completely relevant but redundant in information-theory terms, the present finding of initial interference rather than facilitation under the control-group procedure runs counter to previous results with this variable (e.g., Bourne & Haygood, 1959; Garner, 1962; Haygood & Bourne, 1964). The discrepancy may reflect the complete absence of any initial representation whatever of the added elements under the present successive-addition procedure, whereas with the geometric forms employed in the concept-identification experiments of Bourne and Haygood, such dimensions (e.g., size, color) are held constant at some fixed value but are nonetheless present under the corresponding "nonredundant" procedures. For these and other reasons, greater interitem interference appears to be involved in concept-learning situations (Fallon & Battig, 1964), so that any stimulus redundancy therein may also be more needed to combat this increased interference than under the present rote-learning conditions. Also noteworthy in this connection is an unpublished finding in the Bourne-Haygood experiments, that under conditions of stimulus redundancy *S*s show increased learning of additional stimulus dimensions during the course of the required criterion run of 16 consecutive correct responses.[2] This result is consistent with the recall of added elements observed in the present study, as well as supportive of the present interpretation.

---

[2] L. E. Bourne, Jr., personal communication, 1964.

A more direct test of the proposed explanation for the facilitation produced by the successive addition of stimulus elements would call for the investigation of this variable under conditions of systematic manipulation of the amount of interpair interference within the list, by varying length of list, response similarity, or other nonstimulus factors related to such interference. Also deserving of systematic investigation would be the relative effects of introducing nonoverlapping distinctive elements (as in the present study) as compared with elements common to or shared by one or more of the stimulus terms. Clearly the successive-addition procedure represents a promising technique for the continued investigation of the role played by stimulus cues in paired-associate tasks.

REFERENCES

Archer, E. J. A re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychol. Monogr.*, 1960, **74** (10, Whole No. 497).

Battig, W. F., Brown, S. C., & Nelson, D. Constant vs. varied serial order in paired-associate learning. *Psychol. Rep.*, 1963, **12**, 695–721.

Bourne, L. E., Jr., & Haygood, R. C. The role of stimulus redundancy in concept identification. *J. exp. Psychol.*, 1959, **58**, 232–238.

Brown, S. C. Interpair interference as a function of level of practice in paired-associate learning. *J. exp. Psychol.*, 1964, **67**, 316–323.

Brown, S. C., & Battig, W. F. Partial serial-position constancy in nonserial paired-associate learning. *J. verbal Learn. verbal Behav.*, 1962, **1**, 42–47.

Chang, J.-J., & Shepard, R. N. Meaningfulness in classification learning with pronounceable trigrams. *J. verbal Learn. verbal Behav.*, 1964, **3**, 85–90.

Fallon, D., & Battig, W. F. Role of difficulty in rote and concept learning. *J. exp. Psychol.*, 1964, **68**, 85–88.

Garner, W. R. *Uncertainty and structure as psychological concepts.* New York: Wiley, 1962.

Haygood, R. C., & Bourne, L. E., Jr. Forms of relevant stimulus redundancy in concept identification. *J. exp. Psychol.*, 1964, **67**, 392–397.

Jenkins, J. J. Stimulus "fractionation" in paired-associate learning. *Psychol. Rep.*, 1963, **13**, 409–410.

Johnson, R. B. *Recognition of nonsense shapes as a function of degree of congruence among components of the pretraining task.* (Doctoral dissertation, University of Virginia) Ann Arbor, Mich.: University Microfilms, 1964, No. 64–12, 393.

Newman, S. E. Performance during paired-associate training as a function of number of elements comprising the stimulus term. Technical Report No. 4, 1963, North Carolina State University, Contract Nonr 486(08), Office of Naval Research.

Sundland, D. M., & Wickens, D. D. Context factors in paired-associate learning and recall. *J. exp. Psychol.*, 1962, **63**, 302–306.

Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, 1944.

Underwood, B. J. Stimulus selection in verbal learning. In C. N. Cofer & B. S. Musgrave (Eds.), *Verbal behavior and learning.* New York: McGraw-Hill, 1963.

Underwood, B. J., & Schulz, R. W. *Meaningfulness and verbal learning.* Philadelphia: Lippincott, 1960.

(Received April 16, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 1, 94–97

SOURCES OF TRANSFER FROM ORIGINAL TRAINING TO DISCRIMINATION REVERSAL[1]

W. B. COATE[2] AND R. ALLEN GARDNER[3]

Wellesley College

104 white rats were trained on 2 simultaneous discrimination problems; Problem A was presented on 3 out of 4 trials, Problem B on 1 out of 4 trials. Half of the Ss were reversed on Problem A and half on Problem B. Whether reversed immediately after reaching criterion on Problem A or after overtraining on Problem A, groups that were reversed on Problem A performed significantly poorer than controls that were treated identically except that they were reversed on Problem B instead of Problem A. Also, a source of transfer was identified that could account for a positive effect of overtraining trials on discrimination reversal under some conditions.

Under certain, as yet unspecified, conditions there is a positive relation between the rate of discrimination reversal and the number of postcriterion trials with the original problem. Mackintosh, 1962, 1963; and others cited by Erlebacher, 1963, have found such an effect although Clayton, 1963; Hill and Spear, 1963; Erlebacher, 1963; and others cited by Erlebacher, 1963, have not. It has been suggested that positive transfer from original training to discrimination reversal is based on discriminative processes (S-R habits or cognitive processes) that develop late in original training and make it easier to discriminate S+ from S− during both original training and discrimination reversal.

There are, however, other bases for positive transfer in the overtraining procedure. Typically, the overtrained animals have not only received additional trials with the original problem, but they have also received additional handling, additional adaptation to the deprivation schedule, and additional exposure to the procedure and apparatus of the experiment before the beginning of discrimination reversal. While not strictly speaking, discriminative in nature, these concomitants of overtraining could account for the positive transfer that has been demonstrated so far.

In order to demonstrate that positive transfer from overtraining to discrimination reversal is based on a discriminative process, it is necessary first of all to control for the nondiscriminative concomitants of overtraining. This was the object of the present experiment.

METHOD

Subjects.—The *S*s were 104 male Sprague-Dawley rats obtained from the Charles River Breeding Laboratories when they were between 90 and 120 days old.

Apparatus.—The experimental compartment used for Discrimination Training is shown in Fig. 1. It consisted of a choice chamber and a treadle chamber connected by a narrow passageway and covered by a lid. The front wall of the choice chamber was constructed in two sections separated by a wedge-shaped partition. In the middle of each section, flush with the floor of the chamber, was a 3-in. circular opening. Each opening was covered by a clear Plexiglas panel (P), hinged at the top so that it could be pushed outwards to actuate a microswitch.

---

[1] Supported by Grant MH 4337-02 from the National Institutes of Health.

[2] Now at Hazleton Laboratories, Falls Church, Virginia.

[3] Now at the University of Nevada.

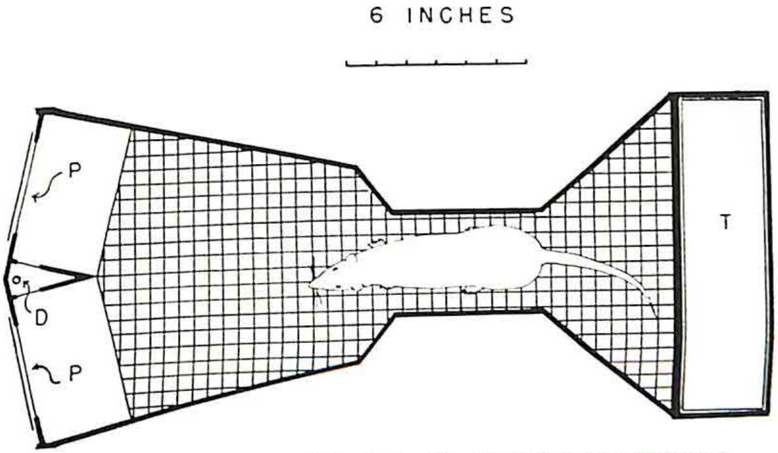


FIG. 1. Diagram of the Discrimination Training compartment.

Outside the chamber, just beyond each panel were the screens of two IEEE projectors (not shown). When the screens were dark, both could be lighted by a momentary depression of the treadle (T), and both were darkened by the first push on either panel. A correct choice was rewarded with a bead of water, delivered automatically by a solenoid-operated dipper, through an aperture (D) in the floor of the choice chamber. Cutouts in the partition permitted *S* to take his rewards from either side.

A second apparatus, used for Preliminary Training, had the same floor plan as the compartment shown in Fig. 1. The front wall of the Preliminary Training compartment could be removed and replaced by a front wall that had only one circular opening and panel in its center. The panels used in Preliminary Training were translucent rather than transparent, and were individually lighted by lamps outside the compartment. When lighted, these panels presented a homogeneous white field.

The experimental rooms were maintained in darkness, with a sound screen of white noise; the automatic control and recording apparatus was kept in a separate room.

Deprivation.—Dry food was available in the living cages at all times. For 6 days before the beginning of Preliminary Training, each *S* was placed on a 23-hr. schedule of water deprivation. During Preliminary Training and Discrimination Training, each *S* was given free access to water in his living cage for 30 min. each day, and was deprived of water for approximately 23 hr. at the start of each daily training session.

Preliminary Training.—Each *S* was handled once daily for 4 days before the beginning of Preliminary Training. Preliminary Training consisted of seven, daily, 30-min. sessions in the Preliminary Training compartment followed by one 48-trial session in the Discrimination Training compartment. During Days 1–4, the single-paneled front wall of the choice chamber was in place. On Day 1, *S* was confined to the choice chamber by a block placed in the passageway. The single panel remained lighted throughout the session and all panel pushes were rewarded. On Days 2–4, the block was removed and *S* was free to move through the passageway between the choice chamber and the treadle chamber. When the panel was pushed while it was lighted, a reward was delivered and the panel was darkened immediately. The panel was relighted when *S* depressed the treadle. On Days 5–7, the single-paneled front wall was replaced by the double-paneled wall with its partition as shown in Fig. 1. Now, when *S* depressed the treadle, one panel was lighted and the other remained dark. Pushing the lighted panel was rewarded, pushing the dark panel was not rewarded, and pushing either panel darkened both. Day 8 of Preliminary Training consisted of 48 trials conducted in the Discrimination Training apparatus under the procedure of Days 5–7.

Discrimination Training.—There were 48 trials on the first day of Discrimination Training and 96 trials on each day after that during both original training and discrimination reversal. The *discriminanda* consisted of 2.5-in. circular fields of stripes. Each stripe was 0.5 in. wide and there were three black and two white stripes. The discriminanda differed only in their angular orientation and the four angles that were used were 0°, 30°, 90°, and 150°. For half of the *S*s, Problem A

TABLE 1

EXPERIMENTAL DESIGN

Group	Criterion Problem	Reversal Problem
AA	A	A
AB	A	B
BA	B	A
BB	B	B

was $0^\circ(+)$ vs. $30^\circ(-)$ and Problem B was $90^\circ(+)$ vs. $150^\circ(-)$, and for the other half, Problem A was $90^\circ(+)$ vs. $150^\circ(-)$ and Problem B was $0^\circ(+)$ vs. $30^\circ(-)$. Thirty degrees and 150° were paired with 0° and 90° on the basis of pilot data which indicated that this would tend to equate the difficulty of the two problems).

During *original training*, the S+ and S− of Problem A were presented together for 36 out of every 48 trials, and the S+ and S− of Problem B were presented for the remaining 12 trials. The 12 Problem B trials were evenly distributed over each block of 48, and left/right orientation of S+ and S− was counterbalanced and evenly distributed over each block of 48 trials for both Problem A and Problem B.

For half of the *S*s original training terminated with the trial on which they reached a criterion of 15 correct out of 16 consecutive Problem A trials, regardless of their performance on Problem B. For the other half, original training terminated with the trial on which they reached a criterion of 15 correct out of 16 consecutive Problem B trials, regardless of their performance on Problem A.

Reversal training began on the day following termination of original training and lasted for 7 days. Only one Problem, either A or B, was presented during all trials of reversal training. Half of the *S*s were reversed on Problem A and the other half on Problem B. The four groups which resulted are shown in Table 1. Twenty-six *S*s were randomly assigned to each group in such a way that all four groups were approximately equated with respect to the time of day and week of the experiment during which data were collected.

RESULTS AND DISCUSSION

The mean number of Problem A trials required to reach criterion on Problem A was 330.0. The results of a 4 × 2 analysis of variance performed on these data were: $F\,(3, 96) = 1.18$ for the four groups, and $F\,(1, 96) = 0.33$ for the two replications—i.e., whether $0^\circ(+)$ vs. $30^\circ(-)$ or $90^\circ(+)$ vs. $150^\circ(-)$ served as Problem A. Neither F approaches significance. A mean of 354.6 postcriterion Problem A trials was received by Groups BA and BB before they reached criterion on Problem B, and t for the difference between Groups BA and BB was 1.14 ($p > .25$).

Performance of all four groups during discrimination reversal is plotted in Fig. 2. Analysis of variance was performed on the total number of correct choices during the 7 days of discrimination reversal. The result was $F\,(3, 100) = 36.2$, $p < .001$. By Duncan's multiple-range test (Edwards, 1960, pp. 136–140) all of the differences between groups shown in Fig. 2 were significant beyond the .01 level of confidence except for the difference between Groups AA and BA which, however, was significant beyond the .05 level.

Groups AA and AB were treated

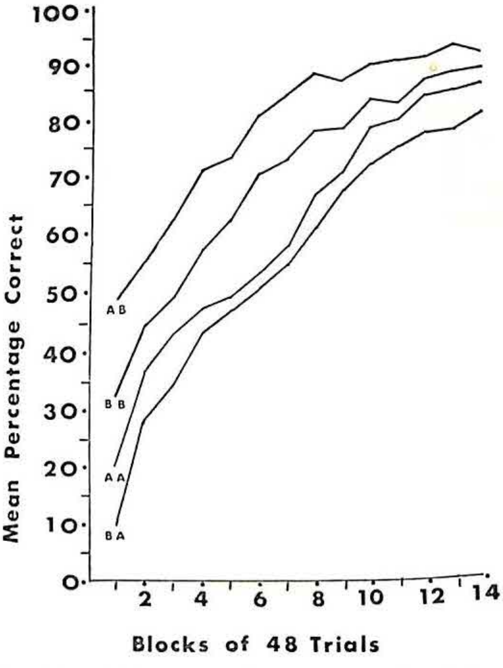


FIG. 2. Performance during reversal training.

identically up to the beginning of reversal training. The inferior performance of Group AA indicates that reversal on Problem A which had been learned to criterion was more difficult than reversal on Problem B which had not been learned to criterion. Groups BA and BB were also treated identically up to the beginning of reversal training. The inferior performance of Group BA indicates that reversal on Problem A which had been presented beyond the point of criterion learning was more difficult than reversal on Problem B which had only been learned to criterion.

Groups AA and BB were reversed on the problem which they had just learned to criterion, and were thus equated for level of acquisition on their reversal problems. The superior performance of Group BB indicates that the net effect of the additional training that they received yielded positive transfer during discrimination reversal. This specifies a possible source of uncontrolled variance that has frequently been overlooked in previous investigations of overtraining and reversal.

The comparison between Groups AA and BA is most similar to those made in previous studies of overtraining and reversal. The inferior performance of Group BA indicates that this experiment must be added to the growing list of those which have found negative transfer from overtraining to discrimination reversal even when nondiscriminative sources of positive transfer were confounded with the overtraining.

All of the differences shown in Fig. 2 were significant, and all are consistent with the conclusion that; both before and after the criterion of original learning was reached, performance during discrimination reversal was (*a*) a positive function of the amount of exposure to the experimental procedure, and (*b*) a negative function of the number of trials on the original problem. These two sources of transfer have been pitted against each other in previous studies of overtraining and reversal, and this might account for the fact that the net effect of overtraining has been positive in some experiments and negative in others.

An incidental finding was the substantial independence of Problem A and Problem B as indicated by the significant difference between Groups AA and AB and between Groups BA and BB. This point was pursued informally by returning eight *S*s of Group BB to the original version of Problem A after they had finished their 7 days of reversal training on Problem B. Seven of these *S*s obtained scores of 91 or more correct out of 96 trials, the eighth obtained a score of 84 out of 96. Clearly, performance on the original version of Problem A could not have been very much impaired by the 672 trials of reversal training with Problem B. This technical feature of the present experiment may be useful in the design of experiments aimed at other problems in discrimination learning.

REFERENCES

Clayton, K. N. Reversal performance by rats following overlearning with and without irrelevant stimuli. *J. exp. Psychol.*, 1963, **66**, 255–260.

Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1960.

Erlebacher, A. Reversal learning in rats as a function of percentage of reinforcement and degree of learning. *J. exp. Psychol.*, 1963, **66**, 84–90.

Hill, W. F., & Spear, N. E. A replication of overlearning and reversal in a T maze. *J. exp. Psychol.*, 1963, **65**, 317.

Mackintosh, N. J. The effects of overtraining on a reversal and a nonreversal shift. *J. comp. physiol. Psychol.*, 1962, **55**, 555–559.

Mackintosh, N. J. Extinction of a discrimination habit as a function of overtraining. *J. comp. physiol. Psychol.*, 1963, **56**, 842–847.

(Received April 17, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 1, 98–101

A TEST OF A DISCRIMINATION HYPOTHESIS OF SECONDARY REINFORCEMENT[1]

NANCY A. MYERS AND JEROME L. MYERS

University of Massachusetts

96 preschool children were assigned to 6 groups which differed with respect to the method of establishing a light as a secondary reinforcer (S^r) in a free-operant situation, and with respect to the schedule of S^r in a choice situation designed to measure S^r strength. Several predictions, generated by a discrimination hypothesis of S^r, were tested and, with one exception, substantiated.

This experiment had two main purposes: (*a*) to evaluate several predictions generated by a discrimination hypothesis of secondary reinforcement (S^r) (Bitterman, Fedderson, & Tyler, 1953); (*b*) to extend the empirical results of studies of choice behavior of preschool children following different schedules of training with a potential secondary reinforcer (PS^r) and primary reward (Fort, 1965).

To accomplish these ends, two variables were manipulated: the method of establishing the secondary reinforcer in a button-pressing situation, and the method of testing S^r effects. With respect to the establishment of S^r, the PS^r appeared either following only those responses that were also primarily reinforced, or following all responses regardless of primary reinforcement. With regard to the test of S^r effects, there were three possible conditions: (*a*) one third of the *S*s chose between a button that always yielded the PS^r and a button that always yielded a novel stimulus; (*b*) another third of the *S*s chose between a button which led to the PS^r on 20% of presses and a novel stimulus button; a final third of the *S*s chose between a button that always yielded the PS^r and a button that led to the PS^r on 20% of presses.

METHOD

Apparatus.—The apparatus consisted of three units—an *S* box, a control box, and a Western Union tape transmitter. The *S* box had mounted vertically on its front three lucite plates, three parallel push buttons, and a slot tray. The plates were placed in a column on the right, and the buttons on the left. The slot tray was located at the bottom center of the box. Sliding Masonite panels permitted any pair of plates and buttons to be concealed from the view of *S*. A red and a green bulb were placed behind each lucite plate. When lighted, each of these illuminated the lucite plate clearly, producing a distinct red or green. One of these lights served as the potential secondary reinforcer (PS^r) for each of the *S*s. Candy-coated chocolates (M & Ms) were used as primary reinforcers and were dispensed to the slot tray through a tube inside the *S* box.

The control box and the Western Union tape transmitter were connected to the *S* box by a long cable, and were kept behind *S* and at the other end of the room, near *E*. The presentation of the appropriate training and testing conditions was accomplished through this automatic programing device.

Subjects.—The *S*s were 48 boys and 48 girls between the ages of 4 yr., 5 mo. and 6 yr., 3 mo., attending nursery schools and kindergartens in Menlo Park and Palo Alto, California.

Procedure.—The *S*s were contacted in their classrooms, and taken to a separate room for experimental sessions. The *E* pressed the center button on the *S* box (the top and bottom ones were concealed), and received a candy. All *Ss* were told that pressing the button would sometimes deliver a candy. In addition, half the *Ss* were told that a colored light would accompany the candy; the other *Ss* were informed that the colored light would occur with each button press. During the subsequent training phase, each *S* was required to respond 75 times; 15 of these responses produced candy on a random schedule. The center button and light were then covered, and the other two revealed. The *E* pressed each of the buttons once to demonstrate that a red light was seen in response to one button, a green light to the other. The child was then instructed to play with the two buttons as long as he wanted.

---

[1] This research was supported by funds from National Institutes of Health Grant MH-02620-04.

The authors wish to thank Nettie Wise, Director, Menlo-Atherton Nursery School, Menlo Park, California, and I. L. Stewart, Director, Educational Guidance Center, Palo Alto, California, for their cooperation in providing facilities and *S*s for this study.

Although *S* was allowed to respond at his own rate throughout the experimental session, any *S* who required more than 8 min. to complete the training phase was eliminated. During the testing period, *E* recorded the number of responses to each button which occurred during each minute, until *S* chose to stop, or until the end of 10 min., at which time *E* terminated the session. Any *S* who stopped before the end of the test received a score of zero for each minute that he missed.

Design.—Six experimental groups of 16 *Ss* (8 boys, 8 girls) each were employed. The groups were equated for mean age and school representation. All groups responded 75 times during conditioning and received 20% candy reinforcement. One half of the *Ss* in each group was trained with red as the PS^r, while the other half was trained with green. One half of each of these groups was presented the PS^r as the top light in the testing period, while the other half found it on the bottom. The groups differed with respect to the scheduling of the PS^r during conditioning and with respect to the stimulus choice available during testing. There were two possible conditioning schedules:

1. PAS (positive-association stimulus)—The PS^r was presented following only those 15 responses which also resulted in the presentation of candy.
2. MAS (mixed-association stimulus)—The PS^r was presented following all 75 responses.

There were three possible test conditions:

1. 100% S^r vs. 100% NS^r—The *Ss* were required to choose between a button that always yielded the colored light previously paired with candy and a button that always yielded a colored light that had not been present during conditioning (NS^r).
2. 20% S^r vs. 100% NS^r—The *Ss* were required to choose between a button that yielded S^r following 20% of the responses and a button that always yielded a colored light that had not been present during conditioning (NS^r).
3. 100% S^r vs. 20% S^r—The *Ss* were required to choose between a button that always yielded S^r and one that yielded S^r following only 20% of the responses.

The combination of conditioning schedules and testing choices resulted in the six experimental groups.

RESULTS

Figure 1 presents the mean number of responses made to each button for each experimental group. Separate analyses of variance were performed on the data of the four groups choosing between S^r and NS^r buttons, and on the data of the two groups choosing between 20% S^r and 100% S^r buttons. The dependent measure was the number of responses made to each button during successive 2-min. periods of time.

Considering first the data for the four groups who chose between an S^r and NS^r button, it is clear that the S^r button was preferred, $F(1, 60) = 28.94$, $p < .001$. Averaging over

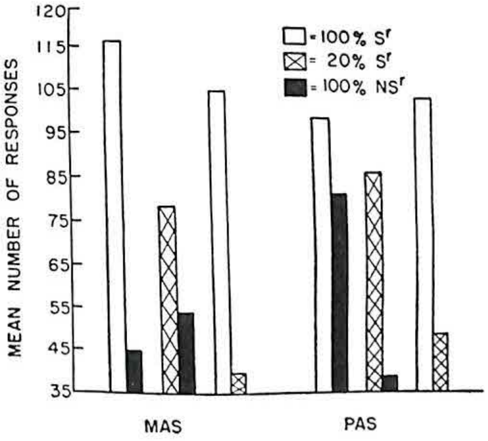


FIG. 1. Mean number of responses on each button for each experimental group.

PAS and MAS training methods, 100% presentation of the S^r for button pressing led to a greater preference for that button over the NS^r button than did 20% presentation, F (1, 60) $= 10.62$, $p < .01$. With regard to the interaction of training methods and percentages, the 100% S^r button was more preferred to the 100% NS^r button by the MAS than by the PAS group, while the 20% S^r button was more preferred to the 100% NS^r button by the PAS than by the MAS group. These findings are reflected in the significant Methods × Percentages × Buttons interaction, F (1, 60) $= 7.38$, $p < .01$.

The mean number of responses to the S^r button dropped from 24.45 in the first 2-min. period of testing to 15.64 in the last 2 min. of testing, while the mean response to the NS^r button dropped from 13.98 to 8.77. These effects of Time and Time × Button were highly significant, $F(4, 240) = 44.58$, $p < .001$; $F(4, 240) = 6.24$, $p < .001$.

For the two groups involved, the 100% S^r button was clearly preferred to the 20% S^r button, F (1, 30) $= 34.42$, $p < .001$. Over time the response rate dropped significantly on both these buttons, F (4, 240) $= 35.98$, $p < .001$.

Discussion

According to a discrimination hypothesis of secondary reinforcement, response strength, during a test in which primary reinforcement is absent, is positively related to the degree of similarity between the patterns of stimulation occurring in conditioning and in the test. The *S* makes the secondarily reinforced response not to gain a "new reward" but because the presence of a stimulus previously present in the conditioning phase has misled him into believing that primary reinforcement will eventually be forthcoming.

The relevance of the discrimination hypothesis to the manipulations of the present study may be clarified by Table 1. In conditioning, MAS groups initially press the middle button (R_m). On all trials, the result is a colored light, the PS^r. They then approach the candy magazine (R_a). On 20% of the trials, candy (S^R) is presented. The distinction between PAS and MAS groups is that, for the PAS *S*s, on 80% of the trials, R_m produces no change in stimulation ($\bar{S}$).

In the test phase, one MAS and one PAS group is required to choose between a button that always yields the previous PS^r (which is now designated as S^r) and a button that always yields a colored light not previously experienced (NS^r). For the MAS group tested in this manner, pressing the top button (R_t) always produces an S-R chain identical to that occurring on 80% of conditioning

TABLE 1

Stimulus-Response Analysis

% Occurrence	
Conditioning for MAS Groups	
20 80	$R_m \rightarrow PS^r \rightarrow R_a \rightarrow S^R$ $R_m \rightarrow PS^r \rightarrow R_a$
Conditioning for PAS Groups	
20 80	$R_m \rightarrow PS^r \rightarrow R_a \rightarrow S^R$ $R_m \rightarrow \bar{S}$
Test for the 100% S^r Groups	
100 100	$R_t \rightarrow S^r \rightarrow R_a$ $R_b \rightarrow NS^r$
Test for the 20% S^r Groups	
20 80 100	$R_t \rightarrow S^r \rightarrow R_a$ $R_t \rightarrow \bar{S}$ $R_b \rightarrow NS^r$
Test for 100% S^r vs. 20% S^r Groups	
20 80 100	$R_t \rightarrow S^r \rightarrow R_a$ $R_t \rightarrow \bar{S}$ $R_b \rightarrow S^r \rightarrow R_a$

trials. For the comparable PAS group, no such identity of conditioning and test chains exists. This S-R analysis leads to the prediction that the MAS group will show a greater preference for the 100% S^r button than the PAS group. The prediction was substantiated by the data.

Next, consider the MAS and PAS group whose test involves a choice between a button that yields the S^r on 20% of the trials and $\bar{S}$ on 80%, and a button that always yields an NS^r. For the PAS group, pressing the top button will, on 80% of these presses, yield a chain identical to that experienced on 80% of conditioning trials. For the MAS group which experiences on only 20% of R_t test responses a chain that occurred on 80% of conditioning trials, generalization, and therefore the S^r effect, should be much less. The prediction, therefore, is that a greater preference for the 20% S^r button will be shown by the PAS group, as was indeed the case.

The S-R model breaks down when the choice between two differentially reinforced buttons is considered. In the choice between 100% S^r and 20% S^r, the discrimination hypothesis predicts that the MAS group should prefer the 100% S^r button; the PAS group should prefer the 20% S^r button. Both groups prefer the 100% S^r button. In fact, the distribution of responses over the two buttons is very similar for the PAS and MAS procedures. The preference for 100% S^r is consistent with previous findings by Myers and Myers (1962, 1963, 1964). Those studies differed from the present one in that *S*s were tested for resistance to extinction on a single button; S^r effects were evaluated on the basis of the difference between groups receiving 100% S^r and groups receiving 20% S^r. The resistance to extinction data also are not sensitive to the PAS-MAS manipulation (Myers & Myers, 1964). Apparently, the discrimination hypothesis accounts for data resulting from a choice between two responses, only one of which produces any S^r. Under other test conditions, the method of conditioning is not relevant and S^r is strongest when its percentage of occurrence is greatest.

REFERENCES

BITTERMAN, M. E., FEDDERSON, W. E., & TYLER, D. W. Secondary reinforcement and the discrimination hypothesis. *Amer. J. Psychol.*, 1953, **66**, 456–464.

FORT, J. G. Learning of a discrimination response in preschool children based on secondary reinforcement established through three training techniques and two schedules of reinforcement. *Child Develpm.*, 1965, in press.

MYERS, N. A., & MYERS, J. L. Effects of secondary reinforcement schedules in extinction on children's responding. *J. exp. Psychol.*, 1962, **64**, 586–588.

MYERS, J. L., & MYERS, N. A. Secondary reinforcement in children as a function of conditioning associations, extinction percentages, and stimulus types. *J. exp. Psychol.*, 1963, **65**, 455–459.

MYERS, J. L., & MYERS, N. A. Secondary reinforcement in children as a function of conditioning associations and extinction percentages. *J. exp. Psychol.*, 1964, **68**, 611–612.

(Received April 18, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 1, 102–108

ACOUSTIC SIMILARITY AND INTRUSION ERRORS IN SHORT-TERM MEMORY[1]

WAYNE A. WICKELGREN
Massachusetts Institute of Technology

36 undergraduates were given lists of 8 items (4 digits and 4 letters) presented at .75 sec/item with ordered recall instructions. Ss were either to copy the items as presented and then recall or to recall the items without prior copying. Recall after copying was slightly worse than recall without copying ($p < .05$). In both conditions, intrusions among letters and between letters and numbers tended to follow acoustic similarity ($p < .001$). Vowel similarities were more important than consonant similarities, but both effects were significant ($p < .001$). There was a correlation of .58 between the number of letters similar to a presented letter and the frequency with which that letter was forgotten ($p < .01$). Short-term memory appears to use an auditory or speech-motor code.

Conrad (1962, 1964) investigated intrusion errors in the immediate recall of lists of six letters drawn from the following 10 letters: B C P T V, F M N S X. He found that letters whose pronunciation ends with an "ē" sound (B C P T V) tend to be confused with each other in recall and letters whose pronunciation begins with an "ĕ" sound (F M N S X) tend to be confused with each other in recall. This was true even though the letters were presented *visually* at the rate of .75 sec/letter under conditions where the probability of perceptual error was known to be negligible.

It is tempting to draw three conclusions from Conrad's study: (*a*) Short-term storage is auditory (or speech-motor). (*b*) Acoustically similar items are represented by similar traces (either overlapping sets of neurons or similar patterns of firing of neurons). (*c*) Partial forgetting of an item is possible, producing intrusion errors that share the unforgotten property common to both the original item and the intrusion.

The present study attempts to provide further support for these conclusions by investigating intrusions involving all 26 letters and the digits 1 through 9. If the above conclusions are valid, there should be a significant tendency for confusions among letters and between letters and digits based on *any* vowel or consonant acoustic similarity. In addition to the ē and ĕ sounds common to many letters and digits, there are other vowel sounds (ā, ū, ī, ō) and consonant sounds (/k/, /s/, /j/, etc.) common to two or more letters or digits. If Conclusion *c* is valid, the frequency of letter intrusions for any letter should correlate with the number of letters which share a sound in common with the presented letter. Similarly, the number of digit intrusions for any letter and the number of letter intrusions for any digit should correlate with the number of digits or letters, respectively, with which a common sound is shared.

The present study differed in several ways from Conrad's experiment, which should test the generality of the relationship between acoustic similarity and intrusions in short-term recall. First, presentation of the items

---

[1] This investigation was supported by Grant MH 07726-01 from the National Institute of Mental Health, Public Health Service.

was auditory rather than visual. Perceptual errors were excluded by having *S*s copy the list as it was presented, cover what they had written, and then recall the list. Only items that were perceived correctly were scored for correct or incorrect recall. Second, lists contained eight items, rather than six items, because the memory span of our *S*s was much greater. Third, lists contained both letters and digits, not just 10 selected letters as in Conrad's experiment.

METHOD

Lists of eight different items, four letters and four digits in random order, were presented on tape at the rate of .75 sec/item. The letters were drawn randomly from the set of all 26 letters, and the digits were drawn randomly from the set 1 through 9. There were two conditions in the experiment: (*a*) Recall and (*b*) Copy + Recall. In the *recall* condition *S*s were instructed to attempt to fill in eight boxes with the correct list of eight items in the correct order, beginning their recall after the entire list of eight items had been presented. In the *copy + recall* condition *S*s were instructed to copy the eight items while they were being presented, and as soon as they had finished copying they were to cover what they had written and attempt to recall the list, as in the *recall* condition. Each *S* received 50 lists under each condition. Thirty-six Massachusetts Institute of Technology undergraduates taking psychology courses participated in the experiment, half receiving the two conditions in one order and half in the other order.

RESULTS

Conrad scored only lists in which there was one and only one error. In the following analysis this criterion was relaxed to include all extralist intrusions where the correct item was not recalled at some other position in the list and those intralist intrusions which occurred in lists with only one error. The presumed purpose for Conrad's selection method and the explicit purpose of the present selection method is to eliminate cases where the intrusion error may be presumed to result from forgetting the exact *position* of an item, rather than partial or complete forgetting of the *item*.

In order to eliminate perceptual errors, the following analysis is limited to those cases where an *S* copied an item correctly and then failed to recall the item correctly. Virtually all of the mistakes in copying (68 out of 84) involved substitution of an acoustically similar item for the presented item, where "acoustic similarity" means "possessing some sound in common." In 70 out of the 84 cases of copying error, there was only one copying error in the list. Substitutions of letters for letters accounted for 58 copying errors; substitutions of numbers for letters accounted for 16 copying errors; substitutions of letters for numbers accounted for 3 copying errors; substitutions of numbers for numbers accounted for 7 copying errors.

Intrusions of Letters for Letters

Table 1 presents the vowel and consonant sounds that occur in two or more English letters. In addition, Table 1 reports the frequency with which letters possessing a common sound are substituted for each other as intrusions in recall after correct copying. Both the vowel and consonant totals are individually significant at far beyond the .001 level in demonstrating a higher frequency of same-sound intrusion errors than would be expected by chance. The vowel sounds have a higher relative frequency of same-sound intrusions to different-sound intrusions than the consonant sounds, but both effects are very significant. The only lingering doubt about the generalization that letter intrusions tend to be acoustically similar to the presented letter is in the case of the consonant similarity classes: B W, L W, D W, and K Q X. The first three classes involve letters which are extremely different in temporal and phonemic length, and the first two classes involve letters in which the common phoneme occupies a very different position in the sequence of phonemes. As an example of both these differences consider B (/biy/) and W (/dubilyuw/). B is two phonemes long, and W is seven phonemes long.[2] /b/ is at the beginning of /biy/, and in the middle of /dubilyuw/.

K (/key/), Q (/kyuw/), and X (/eks/) show no examples of same-sound intrusions out of 13 total intrusions. Each pair of letters possessing the /k/ phoneme does differ in either phonemic length or position of the /k/ phoneme in the sequence of phonemes comprising the pronunciation of the letter. However, the differences are not very large, and differences of this magnitude did not seem to eliminate the greater frequency of same-sound intrusions in letters containing common vowel sounds. The most reasonable explanation for the greater effects obtained with /j/ and /s/ than with /k/ is that the temporal duration of /k/ in pronunciation is much less than the durations of /j/ and /s/ (.08 sec. vs. .18 sec. and .23 sec.—Fletcher, 1953, pp. 62–66). Duration also explains the greater effects with vowels than with consonants. Fletcher (1953) obtained durations of .34 sec. for ē, .30 sec. for ā, and .22 sec. for ĕ, which are much longer than the duration of the average consonant. The differences *among* these vowels in degree of relationship between acoustic similarity and intrusion errors also fit this pattern. Fletcher gives no data on ī and ū, but one suspects they are longer than ĕ, which is *contrary* to the simple hypothesis that the degree of relationship between acoustic similarity and intrusions is dependent on duration of the common sound. The difference between /s/ and /j/ is also contrary to this hypothesis. It is likely that other factors, such as differences in temporal or phonemic length and position of the common sound in the letters possessing a common sound, are also important determinants of the degree of relationship between acoustic similarity and intrusions. We can conclude that acoustic similarity is related to intrusions of letters for letters, but we cannot draw any definite conclusions about the factors that affect the degree of this relationship.

Each of the 26 letters was ranked according to two criteria: (*a*) the total number of letters possessing some sound in common with it and (*b*) the frequency of *all* intrusions for the given letter. The Spearman rank-order correlation between these variables was .70, t (24) = 4.8, $p < .001$, two-tailed. This correlation is composed of: (*a*) a correlation of .14 (*ns*) between the number of letters similar to a given letter and the frequency of similar intrusions per similar letter and (*b*) a correlation of .38, t (24) = 2.0, p = .06, two-tailed, between the number of letters similar to a given letter and the frequency of dissimilar intrusions per dissimilar letter. The correlation between the number of letters similar to a given letter and the frequency of forgetting the letter was .58, t (24) = 3.48, $p < .01$, two-tailed. The frequency of forgetting a letter includes omissions as well as all intrusions. It has already been established that when a letter is not correctly recalled, the intrusions substituted for the letter tend to be acoustically similar to the presented letter. The correlations ob-

---

[2] We shall consider syllabic nuclei, such as /iy/ and /uw/ to be single phonemes.

TABLE 1

Frequency of Same-Sound and Different-Sound Intrusions of Letters for Letters in Recall after Correct Copying

Sound	Letters Containing Sound	Same-Sound Intrusions	Different-Sound Intrusions	p
ā = /ey/	A H J K	9	7	.001[a]
ē = /iy/	B C D E G P T V Z	52	29	.001[a]
ĕ = /e/	F L M N S X	14	16	.001[a]
ī = /ay/	I Y	2	7	.05[b]
ū = /yuw/	Q U W	4	8	.001[a]
Vowel total		81	67	.001[a]
/b/	B W	0	10	
/d/	D W	1	10	
/k/	K Q X	0	13	
/l/	L W	1	10	
/s/	C S X	6	15	.01[a]
/j/	G J	8	8	.001[a]
Consonant total		16	66	.001[a]
Total		97	133	.001[a]

Note.—p = probability of obtaining this frequency of same-sound intrusions by chance (χ^2 for large n, binomia test for small n).
/ / indicates the phonemic representation of the sound.
[a] Two-tailed.
[b] One-tailed.

tained here show that the probability of forgetting a letter and making an intrusion error, either similar or dissimilar, is greater for letters with many similar letters than for letters with few similar letters. A letter that shares a common sound with many other letters is more likely to be forgotten than a letter that shares a common sound with few other letters.

Intrusions between Letters and Digits

The findings for intrusions of letters for digits were identical in all respects to the findings for intrusions of digits for letters. Hence, the data are combined in the following analyses. Table 2 presents four categories of intrusions between letters and digits and the frequency of intrusions in each category. The four categories are: (*a*) Confusions involving a letter which possesses a sound in common with the digit and which is in the spelling of the digit, (*b*) Confusions involving a letter which possesses a sound in common with the digit but which is not in the spelling of the digit, (*c*) Confusions involving a letter which has no sound in common with the digit but which is in the spelling of the digit, and (*d*) Confusions involving a letter which has no sound in common with the digit and which is not in the spelling of the digit. The findings in Table 2 are for intrusions occurring in recall after correct copying, just as in Table 1.

The results are clear. Intrusions between letters and digits tend to be acoustically similar, just as intrusions of letters for letters. Intrusions involving items from two different conceptual categories, letters and digits, tend to follow acoustic similarity just as do intrusions involving items from the same category. The analysis of "silent letters" (letters in the spelling of a digit that are not in the pronunciation of the digit) and the analysis of "invisible sounds" (letters not in the spelling of the digit, but possessing a common sound) suggest that it is the sequence of sounds, rather

TABLE 2

Frequency of Same-Sound and Different-Sound Intrusions between Letters and Digits in Recall after Correct Copying

Digit	Letters with Same Sound, in Spelling		Letters with Same Sound, Not in Spelling		Letters with Different Sound, in Spelling		Letters with Different Sound, Not in Spelling
1	(N)	1	(Y)	0	(OE)	1	25
2	(TW)	1	(UQ)	4	(O)	2	12
3	(RE)	5	(BCDGPTVZ)	17	(H)	0	4
4	(FOR)	9			(U)	0	8
5	(FIV)	6	(Y)	4	(E)	1	10
6	(SX)	4	(CK)	3	(I)	0	7
7	(SVN)	3	(CXFLM)	8	(E)	0	15
8	(HT)	4	(AJK)	7	(EIG)	8	19
9	(NI)	5	(Y)	3	(E)	0	10
Total observed *f*		38		46		12	110
Expected *f*		17.6		20.2		10.6	157.6
p		.001		.001		*ns*	

Note.—p denotes the significance level in a 2×2 χ^2 table comparing the frequency of each column with the frequency in the last column.

than the sequence of letters, which is encoded in short-term memory.

The Spearman rank-order correlation between the number of letters acoustically similar to a given digit and the total frequency of intrusions between letters and digits involving that digit was .30, $t\ (7) = 0.84$, *ns*. The correlation between the number of letters acoustically similar to a given digit and the total frequency of forgetting the digit was .44, $t\ (7) = 1.29$, *ns*. Again the results are the same as for intrusions involving only letters, though not significant because the number of degrees of freedom is smaller.

Intrusions of Digits for Digits

The present experiment confirmed a finding by Conrad (1959) that there are no systematic confusions between digits in short-term recall. This is in perfect agreement with the hypothesis that intrusions follow acoustic similarity since every digit is acoustically very different from every other digit. There are fewer acoustic similarities and more differences among digits than among letters or between digits and letters.

Comparison of Recall after Copying with Recall

The data were analyzed for ordered recall, item recall, and position recall for items. The *S*'s report of an item is correct by an ordered recall criterion if and only if the correct item is recalled in the correct position. The *S*'s report of an item is correct by an item-recall criterion if and only if it appears anywhere in his report of the list in question. Position recall, independent of item recall, is obtained by subtracting the item-recall errors from the ordered-recall errors, reducing the total number of possible errors by the same amount when computing the error rate for position recall.

When the two conditions are compared by any of the three measures of recall, the number of errors in each condition for each *S* is counted and one is subtracted from the other to give a difference score favoring one condition or the other. Superiority of one condition over the other is determined by applying the Wilcoxon matched-pairs signed-ranks test to the 36 difference scores.

Recall after copying is worse than

TABLE 3

Comparison of Recall with Recall after Copying: Number of Intrusions

	Intrusions among Letters		Intrusions between Letters and Digits			
	Similar	Dissimilar	Same Sound, in Spelling	Same Sound, Not in Spelling	Different Sound, in Spelling	Different Sound, Not in Spelling
Recall	98	61	51	54	13	109
Recall minus estimate of copying errors	45	56	43	52	13	100
Recall after copying	92	62	38	46	12	110

Note.—The figures for intrusions among letters in recall after copying in Table 3 are less than the totals in Table 1 because an intrusion may be similar or dissimilar in more than one respect. In Table 1, one intrusion was often counted twice. In Table 3, each intrusion was classified as similar in any respect or dissimilar, and so was counted only once. Expected values were computed differently in the two cases.

recall without prior copying in ordered recall, item recall, and position recall ($p < .05$, two-tailed test, $n = 36$, in all three analyses). The differences are not large, but they are in the expected direction. It is not surprising that copying during presentation reduces the memory span since writing is certainly slower than silent rehearsal and must take some time away from silent rehearsal. What is surprising is that the difference is so small. Even though it is slower than silent rehearsal, writing the items in order may be a particularly useful type of rehearsal when recall also consists of writing the items in order. But regardless of the reason, immediate recall of familiar letters and digits presented at .75 sec/item is not much worse when the items are copied than when the items are not required to be copied during presentation.

The central issue in this comparison of recall with recall after copying is whether the relationship between acoustic similarity and intrusions is any different in the two conditions. Table 3 summarizes the totals for similar and dissimilar intrusions among letters and between letters and digits for recall and recall after copying. Since the totals for recall include some perceptual errors as well as recall errors, the copying errors in the recall-after-copying condition have been taken as an estimate of perceptual errors in the recall condition and subtracted from the totals for the recall condition in the second line of Table 3. This adjustment has little effect on the totals for intrusions between letters and digits, but reduces the frequency of similar recall intrusions among letters by half. However, the relationship between acoustic similarity and intrusions is still significant at beyond the .001 level both among letters and between letters and digits. The relationship may be a little weaker when recall is not preceded by copying, but there is no way to be sure of this since we cannot check the assumption that copying errors provide an estimate of the perceptual error rate when the items need not be copied. One can imagine the estimate might be in error in either direction. It is for this reason that primary emphasis was placed on the analysis of recall intrusions that followed correct copying. Only in recall after correct copying can perceptual errors be eliminated from the recall data.

Discussion

The findings of the present experiment are in complete agreement with the findings of Conrad (1962, 1964) and support the three conclusions mentioned in the introduction: (*a*) Short-term storage uses an auditory (or speech-motor) code. (*b*) Acoustically similar items are represented by similar traces. (*c*) Partial forgetting of an item is possible, producing intrusion errors that share the unforgotten property common to both the original item and the intrusion.

There is a one-to-one correspondence between every sound mentioned in this article and a speech-motor response that produces that sound. Therefore, the present findings imply nothing about the sensory vs. motor nature of the short-term memory trace. Also, these findings do not rule out the possibility that visual-sensory or writing-motor systems play some role in short-term memory. What the present findings do imply is that short-term memory uses either an auditory or a speech-motor code for at least *part* of the trace. No evidence available at present requires one to postulate any replication of short-term traces outside of the auditory or speech-motor systems, and any such replication of the trace using a code not isomorphic to an auditory code would have had a tendency to prevent results such as those obtained in the present experiment. Nevertheless, we cannot be positive that these short-term traces use *only* an auditory or speech-motor code. Incidentally, *very* short-term memory for brief simultaneous visual displays (Sperling, 1960) very likely uses a visual code and is beyond the scope of this article.

According to the auditory or speech-motor theory of short-term memory, the letter J is composed of the consonant /j/ and the vowel /ā/. When J is presented, the internal representatives of both /j/ and /ā/ are activated. When /j/ and /ā/ are activated, J is recalled. When only /j/ is activated, either G or J may be recalled. When only /ā/ is activated, A, H, J, or K may be recalled. When neither /j/ nor /ā/ is activated, any intrusion is equally probable and there may be many omissions. Without the assumption that partial recall of a letter or digit is possible, it seems difficult to account for the present findings.

With the partial-recall assumption, it follows that the greater the number of items sharing a common property with a given item, the greater the probability of making an error when the item is only partially recalled. For example, if one remembers that a letter had an /ā/ in it, one has a .75 chance of guessing incorrectly; if one remembers that a letter had an /ē/ in it, one has a .89 chance of guessing incorrectly. Therefore, /ā/ letters should be "forgotten" less often than /ē/ letters, but the effect should be entirely reflected in an increased probability of making a *similar* intrusion error. The present experiment demonstrated increases in the frequency of *both* similar and dissimilar intrusion errors. In fact, the effect for dissimilar intrusions was larger than the effect for similar intrusions. This result suggests that items with more similar items are more likely to be wholly or partially *forgotten* than are items with fewer similar items.

REFERENCES

Conrad, R. Errors of immediate memory. *Brit. J. Psychol.*, 1959, **50**, 349–358.

Conrad, R. An association between memory errors and errors due to acoustic masking of speech. *Nature*, 1962, **193**, 1314–1315.

Conrad, R. Acoustic confusions in immediate memory. *Brit. J. Psychol.*, 1964, **55**, 75–84.

Fletcher, H. *Speech and hearing in communication.* Princeton: Van Nostrand, 1953.

(Received April 18, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 1, 109–116

THE RELATION OF DRIVE TO FINGER-WITHDRAWAL CONDITIONING[1]

MERRILL F. ELIAS[2]

Purdue University

To explore the relation of drive to performance, Ss scoring high, middle, and low on the Taylor Manifest Anxiety scale were administered a finger-withdrawal task with high- or low-shock UCS. For half the Ss shock always followed the CS (classical conditioning); for half, CRs permitted shock avoidance (avoidance conditioning). In Experiment I performance level was higher for avoidance than for classical conditioning. Under high shock, high-anxiety Ss made fewer CRs than middle- and low-anxiety Ss. This relationship was reversed under low shock. In Experiment II high-anxiety Ss tended to perform the poorest in classical and the best in avoidance conditioning, regardless of shock level. Motion picture analyses revealed that Ss conditioned by the classical procedure made more conditioned irrelevant responses than Ss conditioned by the avoidance procedure. It was concluded that drive increase does not necessarily imply increase of any particular response; the nature of the response and its relation to reinforcement appear to be critical.

In eyeblink-conditioning studies *S*s who have received a strong air puff as the unconditioned stimulus (UCS) have made more conditioned responses (CRs) than *S*s who have received a weak air puff; and, although negative and reverse findings have been reported, numerous investigators have found superior eyeblink conditioning for high-anxiety *S*s (Spence, 1958, 1964). These studies generally support Spence's (1958) hypothesis that performance varies directly with drive as mediated by UCS intensity or when inferred from the Manifest Anxiety scale (*MA* scale). However, there is limited information with respect to the relation of drive to performance in other types of simple conditioning. This investigation was undertaken to extend the knowledge of drive effect on performance by exploring the relation of anxiety level and UCS intensity to finger-withdrawal conditioning.

EXPERIMENT I

Two conditioning procedures were used: classical and avoidance. It was hypothesized that (*a*) *S*s who received high-shock UCS would be superior in performance to *S*s who received low-shock UCS in both avoidance and classical conditioning; (*b*) avoidance conditioning would be superior to classical conditioning since a CR served to avoid shock; and, (*c*) classical conditioning would provide the clearest relation between *MA* scale measured drive and performance. In classical conditioning the incidence of undetected voluntary responding, a

---

[1] This article is based on a dissertation which was submitted to the faculty of Purdue University in August 1963 in partial fulfillment of requirements for the PhD. Detailed description of methodology, particularly with respect to characteristics of the electric shock stimuli and motion picture analyses, may be obtained by requesting a microfilm copy of the dissertation through Purdue University interlibrary loan service.

The author wishes to gratefully acknowledge the painfully exacting and, consequently, invaluable assistance of Kenneth M. Michels and Mark W. Stephens in the development of this research.

[2] Now at Rome Air Development Center (EMEDI), Griffiss Air Force Base, New York.

phenomenon which the Iowa group (Spence, 1964) has shown to obscure the positive relation between anxiety level and performance, should be minimal since neither voluntary responses nor CRs permitted shock avoidance. Further, Spence (1958) has suggested that classical conditioning affords the clearest control of higher-order habit variables and is preferred when interest is in isolation of the relation between performance and drive per se.

Method

Subjects.—One hundred and twenty Air Force ROTC cadets who ranged in age from 17 to 20 yr. served as *S*s.

Apparatus.—The *S*s were seated in a normally illuminated room with their right forearm and wrist strapped to the arm of a classroom chair. Normal pressure of the extended middle finger depressed a sensitive microswitch. The conditioned stimulus (CS), a buzzing sound of approximately 61 db. SPL, was generated by an inductorium wired in parallel with a 6-v. dc power source. The UCS (electric shock) was generated by an identically wired inductorium. The CS was presented .4 sec. before the UCS and terminated with the end of the UCS which lasted for .2 sec. Latency of finger withdrawal was recorded by a Standard Electric timer which was wired in series with the microswitch. The timer, activated with the onset of the CS, was stopped by a finger withdrawal of sufficient magnitude to release tension on the microswitch. Finger withdrawal was considered a CR if it occurred within the .4-sec. CS-UCS interval on a trial in which the shock UCS was not presented.

For the avoidance procedure a silver electrode (1 cm. diameter) was placed on the tip of the microswitch and another silver electrode ($3 \times 1\frac{3}{4}$ cm.) was taped to the palmar surface of *S*'s forefinger; thus, withdrawal of the middle finger within the CS-UCS interval permitted avoidance of shock. For the classical procedure both electrodes were taped to *S*'s fingers; the small electrode was taped to the middle finger and the larger electrode was taped to the forefinger. Through this arrangement *S* in the classical procedure could not escape or avoid shock even though he responded within the CS-UCS interval. The CS-UCS interval, time between trials, and UCS duration were controlled by means of five Hunter timers. A 40-db. SPL, 200-cps tone was delivered to *S*'s left ear via earphone to mask the sound of recording equipment.

Procedure.—The *S*s were told that the purpose of the experiment was to test their ability to concentrate under stress. They were asked to count the number of times they heard a buzzing sound (CS) and were told that they would receive shock (UCS) from time to time to make counting difficult. The *S*s were instructed neither to voluntarily remove their fingers from the microswitch nor to resist natural movement of the finger. Preliminary tests of CS neutrality revealed no evidence of pseudoconditioning. After these preliminary tests, five blocks of 50 trials each were presented for a total of 250 trials. Ten of the 50 trials per block were randomly interjected *test* trials in which the UCS was not presented.

Design.—A $3 \times 2 \times 2 \times 5$ factorial design was used. Equal numbers of *S*s were assigned to high-, middle-, and low-anxiety groups on the basis of *MA* scale scores. High-anxiety *S*s had scores of 16 and above, middle-anxiety, 9–15, and low-anxiety, 8 and below. These scores represented the low, middle, and upper thirds of a distribution of scores obtained from 215 Air Force ROTC cadets. The *S*s in each anxiety group were randomly assigned to experimental conditions: half were conditioned by the classical procedure and half by the avoidance procedure. Half of the *S*s in each conditioning group received low (30–70 μamp.) and half received high (75–125 μamp.) shock as the UCS. The Block factor consisted of five levels. Each level represented a block of 10 test trials.

Results

Analyses of variance, summarized in Table 1, will be discussed in relation to Fig. 1. The significant ($p < .01$) superiority of avoidance conditioning is apparent in Fig. 1. The significant ($p < .01$) Blocks main effect reflects the progressive improvement of all *S*s throughout conditioning. The *S*s who received high-shock UCS made significantly ($p < .01$) more CRs than *S*s who received low shock. The significant ($p < .01$) Blocks × Shock interaction reflects the tendency for block-to-block improvement to be greater for the high-shock group.

TABLE 1

Analyses of Variance for Exp. I

Source	*df*	*MS*	*F*
Between *S*s	119		
Conditioning (C)	1	1,700.1656	53.73**
Shock (S)	1	1,225.7056	39.68**
Anxiety (A)	2	13.2712	.42
C × S	1	125.1278	3.95*
C × A	2	18.3022	.58
S × A	2	112.3622	3.56*
C × A × S	2	29.8961	.95
S within-group error	108	31.6455	
Within *S*s	480		
Blocks (B)	4	234.1806	87.94**
B × C	4	7.5461	2.80*
B × S	4	10.8861	4.09**
B × A	8	28.0905	10.52**
B × C × S	4	26.8806	10.09**
B × C × A	8	4.5075	1.69
B × S × A	8	3.3470	1.30
B × C × A × S	8	8.1322	3.05**
B × *S* within-group error	432	2.6692	
Simple effects of shock at:			
Classical conditioning	1	294.0300	9.29**
Avoidance conditioning	1	1,086.8034	34.34**
S within-group error	108	31.6455	

* $p < .05$.
** $p < .01$.

There was a significant ($p < .05$) Conditioning × Shock interaction. Differences between the high- and low-shock groups were greater for avoidance than for classical conditioning; however, analyses of simple effects (Table 1) indicated a significant superiority for the high-shock groups under both classical ($p < .01$) and avoidance ($p < .01$) conditioning.

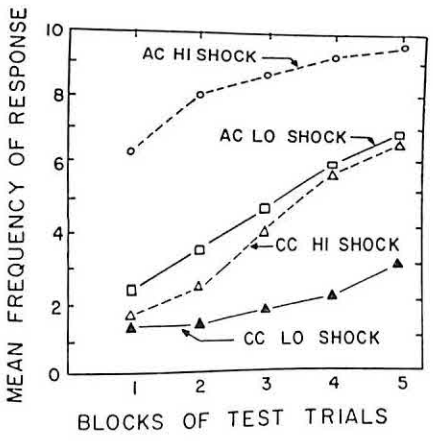


Fig. 1. Performance of classical and avoidance conditioning groups under high- and low-shock UCS for five blocks of test trials averaged over the three levels of anxiety.

The significant ($p < .01$) Blocks × Conditioning × Shock interaction reflects the tendency for differences between shock groups to be greater under avoidance than under classical conditioning for the first three blocks of trials, but for the last two blocks the difference between high- and low-shock groups was slightly greater for classical conditioning.

There were no significant differences among Anxiety groups when scores were averaged over all other experimental conditions and a significant Conditioning × Anxiety interaction was not obtained. The significant ($p < .01$) Blocks × Anxiety interaction may be described as a variation in improvement from block to block for the three Anxiety groups; however, tests of simple effects revealed no significant differences among Anxiety groups at any block. There was a significant ($p < .05$) Shock × Anxiety interaction. Under the low-shock condition mean frequency of CR per block was greater for the high-anxiety group (4.57) than for the middle- (3.40) and low- (2.42) anxiety groups. Under high shock, the high-anxiety group actually made fewer responses (5.40) than the middle- (6.61) and low- (6.61) anxiety groups. Although the Conditioning × Anxiety × Shock interaction was not significant, the largest portion of this Anxiety × Shock interaction may have been due to the poor performance of the high-anxiety group under high-shock classical conditioning; for low-shock classical conditioning mean frequency of CR per block was a monotonic function of increasing anxiety level (1.40, 1.58, 3.98), but for high-shock classical conditioning this relationship was reversed. Mean frequency of CR for the low-anxiety group was 4.94; middle-anxiety group, 4.56; and high-anxiety group, 2.82.

The significant ($p < .01$) Blocks × Conditioning × Anxiety × Shock interaction may be described as a progressive increase in frequency of response from the first to the last block of trials with a varying rate of improvement among and within groups.

Discussion

The superior performance of *S*s who received high-shock UCS in both avoidance and classical conditioning is consistent with hypotheses derived from Spence's model.

In contrast to findings in the present study, Logan (1951) found better classical than avoidance conditioning of the eyeblink response in humans when shock to the corner of the eye was used as the UCS. Among others, Kimble[3] has suggested that these conflicting results might possibly be related to the voluntary nature of finger-withdrawal responses. Although *S*s in the present experiment were instructed not to voluntarily respond and were eliminated from the sample if conditioned responding showed an immediate rise to 100% after the initial response, the possibility of completely involuntary responding cannot be defended. The possible presence of a greater number of undetected involuntary responders in the avoidance group does not vitiate conclusions with respect to the superiority of avoidance conditioning, but it may provide one explanation of the failure to obtain significant differences among Anxiety groups.

Spence (1964) has presented evidence that failures to obtain a positive relation between anxiety and eyeblink-conditioning performance were associated with failure to identify and eliminate voluntary responders. He has pointed out that an inproportionately high number of voluntary responders in a low-anxiety group could result in negative or reverse findings and that voluntary responding seems to increase with increased UCS intensity. This phenomenon could explain the poor performance of high-anxiety *S*s under high shock averaged over the two conditioning procedures. However, the poor performance of the high-anxiety *S*s under high-shock classical conditioning is puzzling. It would be expected that the incidence of voluntary responding under this condition would be negligible since a response was not rewarded by shock avoidance. It was, in fact, consistently followed by shock. Unsystematic observation of responses other than finger withdrawal suggested an explanation.

Experiment II

To determine whether these irrelevant responses could explain the results obtained, the first experiment was repeated and motion pictures were taken of all behavior during conditioning.

Method

The procedure, design, and apparatus were identical to that of the first experiment with the following modifications. A Bollex Paillard 16-mm. motion picture camera with a wide angle lens was hidden from view and used to record all responses other than finger retraction. Camera noise necessitated an increase in the masking tone from the 40-db. SPL of Exp. I to 50 db. SPL. CS intensity was increased correspondingly by replacing the induction coil of Exp. I with a doorbell buzzer of 72 db. SPL. Only 27 *S*s were used; 2 from each anxiety level were randomly assigned to the four experimental conditions: high- and low-shock classical and avoidance conditioning. One *S* from each anxiety level was assigned to a control condition in which shock UCS was never given.

Motion picture analyses.—The *E* and a trained *O* reviewed films for 10 *S*s selected at random from the 27 *S*s used in the experiment. Responses were categorized by muscle group and direction of movement. Following this preliminary analysis, films for all *S*s were analyzed by *E* using as response criteria only those response categories for which there was acceptable *E-O* agreement. A Pearson product-moment correlation coefficient of .9

---

[3] G. A. Kimble, personal communication, January 2 and 23, 1963.

TABLE 2

NUMBER OF *S*s WHO MADE IRRELEVANT RESPONSES

Experimental Group ($N = 24$)	N[a]	Mean Frequency[b]	
		UCR	CR
Flinching—Forward or rearward movement of the head and/or shoulders	21	39.71	4.91
Rearing—Exaggerated forward or rearward head movement accompanied by shoulder hunching	18	22.50	7.78
Stepping—Lifting of toes or heels from the floor	10	7.10	1.10
Grimace—Closing of eyes and facial contortion	8	25.58	14.62
Leg Flexion—Lifting of leg(s) or sliding of foot forward along the floor	8	4.63	1.13
Hand Flexion—Lifting of hand of nonstimulated fingers	4	1.24	0.00
Arm Flexion—Lifting of arm of nonstimulated fingers	2	5.00	0.00
Control Group ($N = 3$)	N[a]	Mean Response Frequency[b]	
All responses combined	2	4.5	

[a] Number of *S*s who made at least one response.
[b] Average number of responses for *S*s who made at least one.

or above between the number of times for each block of trials that a particular response was observed by *E* and *O* was used as the criterion of acceptable agreement.

Results

It may be seen in Table 2 that a variety of responses was elicited by the UCS and, with the exception of hand and arm movements, did become conditioned. The few responses made by the control group suggests that responding in the experimental groups was not due to naturally occurring bodily movements. Since not all *S*s made the same kinds of responses analyses of variance, summarized in Table 3, were performed for two dependent variables: conditioned finger withdrawal and conditioned irrelevant response. A conditioned irrelevant response was defined as the occurrence, on a test trial, of one or more of the responses summarized in Table 2. Figure 2 compares the performance of classical and avoidance groups. As in Exp. I, mean frequency of conditioned finger with-

TABLE 3

ANALYSES OF VARIANCE OF FINGER-WITHDRAWAL AND IRRELEVANT RESPONSE DATA

Source	*df*	Finger Withdrawal		Irrelevant Response	
		MS	*F*	*MS*	*F*
Between *S*s	23				
Conditioning (C)	1	313.6333	8.20**	23.4084	.94
S within-group error	22	38.2470		24.9356	
Within *S*s	96				
Blocks (B)	4	60.9792	14.95**	21.1375	4.58**
B × C	4	6.6959	1.64	15.7208	3.41*
B × *S* within-group error	88	4.0789		4.6155	

Note.—For irrelevant response, simple effects of conditioning at Block 5 were significant—$MS = 80.00$; $F(1) = 9.22$; $p < .01$—when tested with pooled within-cell error ($df = 110$; $MS = 8.6795$) as the denominator.
* $p < .05$.
** $p < .01$.

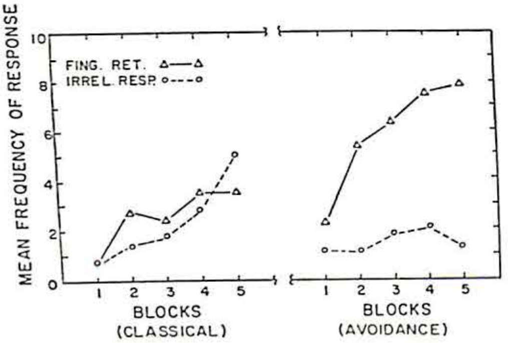


FIG. 2. Performance of classical and avoidance conditioning groups for five blocks of test trials averaged over all other experimental conditions.

drawal was significantly ($p < .01$) higher in avoidance than in classical conditioning. Frequency of conditioned irrelevant response continued to increase throughout the five blocks of classical conditioning, whereas in avoidance conditioning irrelevant responding was extinguishing while finger withdrawal became more frequent. There was, in fact, a significant ($p < .01$) Blocks × Conditioning interaction for irrelevant responding; and, analyses of simple effects, summarized in Table 3, indicated that on the fifth block of trials the classical group made significantly ($p < .01$) more irrelevant responses than did the avoidance group.

The specific purpose of the second experiment was to investigate the relation of shock and anxiety level to performance in each type of conditioning; consequently, separate analyses of variance for classical and avoidance conditioning were performed and are summarized in Table 4. There were no significant differences between either high- and low-shock classical or high- and low-shock avoidance groups for either response criteria. There was a significant Blocks main effect and Blocks × Shock interaction for both finger withdrawal ($p < .05$) and irrelevant responding ($p < .01$) for classical conditioning; i.e., there was a block to block increase in frequency of both irrelevant and finger-withdrawal responding for the high-shock group, but no trend toward either increased

TABLE 4

SEPARATE ANALYSES OF VARIANCE FOR CLASSICAL AND AVOIDANCE CONDITIONING

Source	*df*	Classical				Avoidance			
		Finger CR		Irrelevant CR		Finger CR		Irrelevant CR	
		MS	*F*	*MS*	*F*	*MS*	*F*	*MS*	*F*
Between *S*s	35								
Anxiety (A)	2	12.5165	.76	3.9556	.26	3.3550	.05	7.6900	.75
Shock (S)	1	.1399	.01	12.8000	.81	4.0500	.07	2.4889	.24
A × S	2	2.9055	.18	7.8667	.50	28.8667	.47	7.5983	.73
S within-group error	30	16.4789		15.7289		61.5189		10.2790	
Within *S*s	144								
Blocks (B)	4	4.2722	3.27**	11.4861	4.66**	19.1195	7.49**	.6875	.63
B × A	8	1.2181	.93	1.0969	.45	3.4528	1.35	2.2437	2.05*
B × S	4	3.4445	2.64*	8.6195	3.50**	.3972	.16	.4178	.38
B × A × S	8	.7636	.58	.5195	.21	1.7556	.69	.7761	.71
B × *S* within-group error	120	1.3067		2.4622		2.5522		1.0928	

* $p < .05$.
** $p < .01$.

or decreased responding for the low-shock group.

For avoidance conditioning there was a significant ($p < .01$) Blocks main effect for finger withdrawal reflecting the progressive increase in frequency of responding from block to block. For irrelevant responding there was a significant ($p < .05$) Blocks × Anxiety interaction in avoidance conditioning which may be described as a tendency for irrelevant responding to quickly extinguish for low-anxiety *S*s in contrast to a progressive increase in irrelevant responding for middle- and high-anxiety *S*s.

The inverse relationship between anxiety level and performance found in Exp. I under high-shock classical conditioning was not replicated; however, Fig. 3 shows two interesting relationships between anxiety level and type of conditioning which may partially explain poor performance of high-anxiety *S*s in classical conditioning. (*a*) For avoidance conditioning an increase in anxiety level was associated with an increase in frequency of finger withdrawal. For classical conditioning anxiety level and frequency of finger withdrawal were inversely related. (*b*) The best performers in classical conditioning, the low-anxiety *S*s, made the fewest irrelevant responses while the poorest performers, the high-anxiety *S*s, made more irrelevant than finger-withdrawal responses. On the other hand, in avoidance conditioning, not only was an increase in frequency of finger withdrawal from low- to high-anxiety conditions accompanied by a corresponding increase in irrelevant responding, but at every point on the anxiety continuum frequency of irrelevant responding was considerably less than frequency of finger withdrawal.

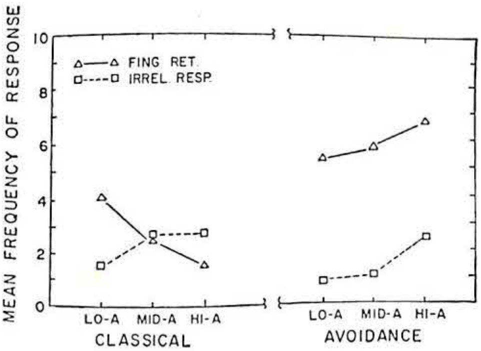


FIG. 3. Performance of high-, middle-, and low-anxiety groups under classical and avoidance conditioning averaged over high- and low-shock conditions.

Discussion

These data suggest that the function of drive is to energize general defense responses, while UCS avoidance acts to selectively reinforce a particular response. In classical conditioning, where shock could not be avoided, finger withdrawal and irrelevant responses were acquired at about the same rate and reached the same level. In avoidance conditioning, where shock was avoided by a finger withdrawal, the finger-withdrawal response showed a progressive acquisition throughout conditioning as irrelevant responding extinguished.

Interference of irrelevant responses with the conditioning of finger withdrawal may explain the relatively poor performance of the classical conditioning group, particularly under high shock-high anxiety in the first experiment and high anxiety in the second experiment where drive probably energized both finger-withdrawal and irrelevant responses. Sheffield (1948) has presented evidence that the inferior conditioning of the running response in Guinea pigs who could not avoid shock was due to the interference of conditioned emotional responses such as hopping, turning, and scrambling. In 1952 Gibson reported that kids conditioned by an avoidance procedure developed a smooth efficient leg flexion response while kids conditioned by a classical procedure performed at a lower level and made as many as 14 defense reactions which may have interfered with conditioning of the leg-flexion response.

Failure, in avoidance conditioning, to obtain significant differences among Anxiety groups was probably due to the extremely low number of *S*s in these groups. Similarly, the likelihood of obtaining undetected voluntary responders in these groups may have been reduced. This might account for the trend toward a positive relation between Anxiety level and avoidance conditioning performance in Exp. II.

These findings do not disconfirm Spence's predictions with regard to drive and performance. They suggest that even in simple forms of learning, increased drive does not necessarily imply an increase in any particular response chosen as the dependent variable. The relation of drive to a variety of responses should be investigated. There is experimental evidence that a distinction between voluntary and involuntary finger-withdrawal responses is not impossible (Moyer & Lindley, 1961; Wickens, 1939; Yacorzynski & Guthrie, 1937). Development of valid criteria for making this distinction may be requisite to a clearer determination of drive effect on finger-withdrawal conditioning.

REFERENCES

GIBSON, E. J. The role of shock in reinforcement. *J. comp. physiol. Psychol.*, 1952, **45**, 18–31.

LOGAN, F. A. A comparison of avoidance and nonavoidance eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 391–393.

MOYER, K. E., & LINDLEY, R. H. Effects of instructions on the extinction of a conditioned finger-withdrawal response. *J. exp. Psychol.*, 1961, **61**, 82–88.

SHEFFIELD, F. D. Avoidance training and the contiguity principle. *J. comp. physiol. Psychol.*, 1948, **41**, 165–177.

SPENCE, K. W. A theory of emotionally based drive (*D*) and its relation to performance in simple learning situations. *Amer. Psychologist*, 1958, **13**, 131–141.

SPENCE, K. W. Anxiety (drive) level and performance in eyelid conditioning. *Psychol. Bull.*, 1964, **61**, 129–139.

WICKENS, D. D. A study of voluntary and involuntary finger conditioning. *J. exp. Psychol.*, 1939, **25**, 127–140.

YACORZYNSKI, G. K., & GUTHRIE, E. R. A comparative study of involuntary and voluntary conditioned responses. *J. gen. Psychol.*, 1937, **16**, 235–257.

(Received April 20, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 1, 117–121

SENSITIVITY OF SUBJECTIVE PROBABILITY REVISION[1]

CAMERON R. PETERSON[2] AND ALAN J. MILLER

University of Colorado

Ss were required to revise their subjective probabilities of hypotheses as the result of information provided by a datum. Subjective probability revisions were then compared with the corresponding correct revisions as calculated via Bayes' theorem. 2 independent variables, prior odds in favor of a hypothesis and the theoretical impact of the datum, both influenced 2 dependent variables, amount and accuracy of subjective probability revision. These effects were such that variations in subjective probability revision were in the same direction, but of smaller magnitude, than corresponding variations in Bayesian probability change.

Several experiments have compared the "intuitive" revision of subjective probabilities with the correct revision as calculated by means of Bayes' theorem. The typical finding is the *S*s are conservative; the magnitude of subjective revision is less than the correct amount (Edwards, Lindman, & Phillips, 1965; Edwards & Phillips, 1964; Peterson, Schneider, & Miller, 1965).

Bayes' theorem specifies the revision of the probability of Hypothesis H as the result of information provided by the occurrence of Datum D (Edwards, Lindman, & Savage, 1963); it may be written as

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)} \quad [1]$$

The $P(H)$ is revised to $P(H|D)$ as a result of the occurrence of D. An equation for two hypotheses, H_a and H_b, which is directly relevant to the present experiment, is obtained by dividing Equation 1 for H_a by Equation 1 for H_b:

$$\frac{P(H_a|D)}{P(H_b|D)} = \frac{P(D|H_a)}{P(D|H_b)}\frac{P(H_a)}{P(H_b)} \quad [2]$$

or, written more simply,

$$\Omega_1 = L\Omega_0 \quad [3]$$

The Ω_0 refers to the odds in favor of H_a over H_b prior to the occurrence of D. The Ω_1 refers to the corresponding odds posterior to the occurrence of D. The L refers to the likelihood ratio; it is the probability of D given H_a divided by the probability of D given H_b. The odds in favor of H_a over H_b are revised from Ω_0 to Ω_1 as a result of the occurrence of D. Thus, L is a measure of the impact of D where impact refers to the amount of revision from Ω_0 to Ω_1.

According to Equation 3 the two variables, Ω_0 and L, determine the correct revision in odds and therefore the correct revision in probabilities. The present experiment manipulated both Ω_0 and L and measured the resulting effect upon the revision of subjective probabilities; it evaluated degree of conservatism as a function of Ω_0 and L.

---

[1] The research reported here was undertaken in the Behavior Research Laboratory, Institute of Behavioral Science, University of Colorado, and is Publication No. 43 of the Institute. The research was supported by a research grant (M-4977) from the National Institute of Mental Health. The authors are indebted to W. Edwards, K. R. Hammond, and F. J. Todd for helpful suggestions and to Dorothy Boucher for assistance in the conduct of the experiment.

[2] Now at the University of Michigan.

METHOD

Experimental design.—The experimental problem required the manipulation of Ω_0 and L, and a measurement of the resulting effects upon subjective probability revision. On each trial, two different kinds of dice served as H_a and H_b. The dice were distinguished by different numbers of black vs. white sides. The Ω_0 was manipulated by drawing either H_a or H_b from an urn containing varying proportions of each. The L was manipulated by varying the ratio of black sides on H_a to black sides on H_b.

The Ω_0's took on the nine different values corresponding to $P(H_a)$s of .1, .2, .3, .4, .5, .6, .7, .8, and .9. These values covered most of the uncertain range for Ω_0. The four levels of L, designated L I, L II, L III, and L IV, were manipulated by assigning the following ratios of black sides on H_a to black sides on H_b—3:2, 4:2, 5:2, and 5:1, respectively. Accordingly, the corresponding ratios of white sides were 3:4, 2:4, 1:4, and 1:5. These four levels range from low impact of D to high impact. A factorial design permitted all combinations of low to high data impact with low to high prior odds.

One measure of subjective probability revision, ΔP, is the absolute difference between the subjective probabilities associated with $P(H_a)$ and with $P(H_a|D)$. The second measure, the *accuracy ratio*, relates subjective revision to Bayesian revision. The numerator of the accuracy ratio is the log likelihood ratio based upon subjective probability revision and the denominator is the log likelihood ratio based on Bayesian revision. The accuracy ratio is 1.0 when subjective revision equals Bayesian revision and it increases with the ratio of subjective revision to Bayesian revision. (The accuracy ratio is described in Peterson et al., 1965).

Subjects.—Forty-two volunteer students from an introductory psychology course served as *S*s in groups of approximately four each.

Apparatus.—Five distinct groups of nine dice each were associated with five different *H*s. Each six-sided die in a group had either one, two, three, four, or five black sides. The remaining sides were white. Five different placards, representing each *H*, illustrated the numbers of black sides and white sides of the relevant *H*s to *S*s.

The response apparatus was an indexed bar with a sliding marker. The location of the marker partitioned the bar into two lengths which represented *S*s' probability estimates for H_a and H_b (this apparatus is described in Peterson & Ulehla, 1964). These estimates were used as measures of *S*'s subjective probabilities.

Procedure.—The experiment consisted of five blocks of 36 trials each. Each trial represented one of the 36 combinations of nine Ω_0's and four *L*s. These combinations were counterbalanced across blocks within the constraint that all Ω_0's associated with a single L were presented consecutively in either ascending or descending order.

Each trial included only two of the *H*s, or two kinds of dice. The die with the greater number of black sides is termed H_a; the die with the greater number of white sides is H_b. The E manipulated Ω_0 by placing 10 dice into an opaque urn with varying proportions of H_a and H_b. For example, seven H_a's and three H_b's yielded an Ω_0 of 7:3. The E then drew one of the 10 dice at random and required S to set prior subjective probability estimates for H_a and H_b to correspond to Ω_0. The E rolled the die from the cup and informed S which side turned up. On the basis of this information, S revised his subjective probability estimates about which die (H_a or H_b) had been rolled, and recorded his revision. This revision then served as the basis for each dependent variable.

RESULTS AND DISCUSSION

According to Table 1, Ω_0, L, and the $\Omega_0 \times L$ interaction all had significant ($p < .001$) effects upon both ΔP and the accuracy ratio. The variable, Blocks, did not in itself have a significant effect, but in some cases interacted significantly ($p < .01$) with Ω_0 and/or L.

Figure 1 illustrates the nature of the effects of Ω_0 and of L upon the accuracy ratio. The accuracy ratio generally increased both as Ω_0 departed from 1.0 and L approached 1.0. On the other hand, the expected probability revision in the Bayesian denominator of the accuracy ratio *decreased* as Ω_0 departed from 1.0 and as L approached 1.0. This reciprocal relation between the accuracy ratio

TABLE 1

Analyses of Variance for Accuracy Ratio and ΔP

Source	df	Accuracy Ratio		ΔP	
		MS	F	MS	F
Prior odds (Ω_0)	8/328	55.34	29.16***	.239	16.11***
Likelihood ratio (L)	3/123	419.69	90.74***	.413	20.82***
Blocks (B)	4/164	1.49	<1	.024	<1
$\Omega_0 \times L$	24/984	5.42	4.79***	.027	4.56***
$\Omega_0 \times$ B	32/1312	1.11	1.49	.017	2.67***
$L \times$ B	12/492	2.36	1.84*	.016	2.10*
$\Omega_0 \times$ B $\times L$	96/3936	1.16	1.62**	.011	1.75***

* $p < .05$.
** $p < .01$.
*** $p < .001$.

and its denominator indicated that variation in the subjective numerator is more sluggish than variation in the Bayesian denominator.

Figure 2 supports this conclusion. With respect to both Ω_0 and L, the subjective ΔP usually varied in the same direction as the corresponding Bayesian ΔP but the magnitude of the subjective variation was less. The subjective graphs are similar to, but evidence less vertical movement than the Bayesian graphs. Vertical compression applied to the Bayesian graphs would result in rough approximations to the subjective graphs. The *S*s were only partially sensitive to variations in Bayesian probability revision. As the difference between Bayesian prior and posterior probabilities increased, conservatism increased. In fact, when the difference between Bayesian priors and posteriors became small enough, such as for Cond. *L* I, *S*s actually revised their subjective probability estimates more than the amount specified by Bayes' theorem.

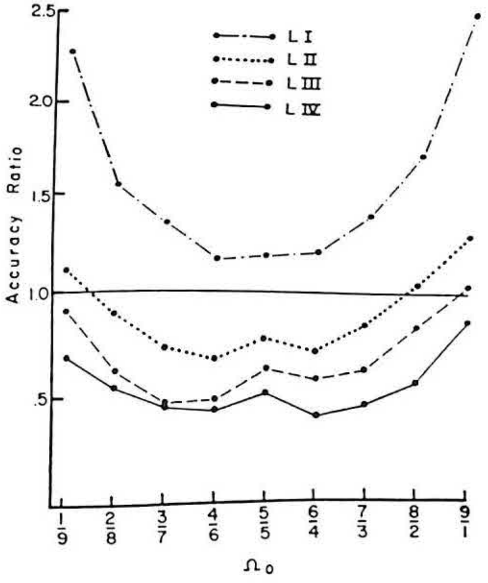


Fig. 1. Accuracy ratio as a function of Ω_0 and L. (The solid horizontal line represents perfect accuracy.)

This partial sensitivity to variation in Bayesian revision was somewhat absent during the earlier blocks but became increasingly evident as the blocks progressed. As *S*s became familiar with the task, they must have become able to detect which conditions called for more movement and which ones for less movement. They then adjusted their responses to eliminate inconsistencies.

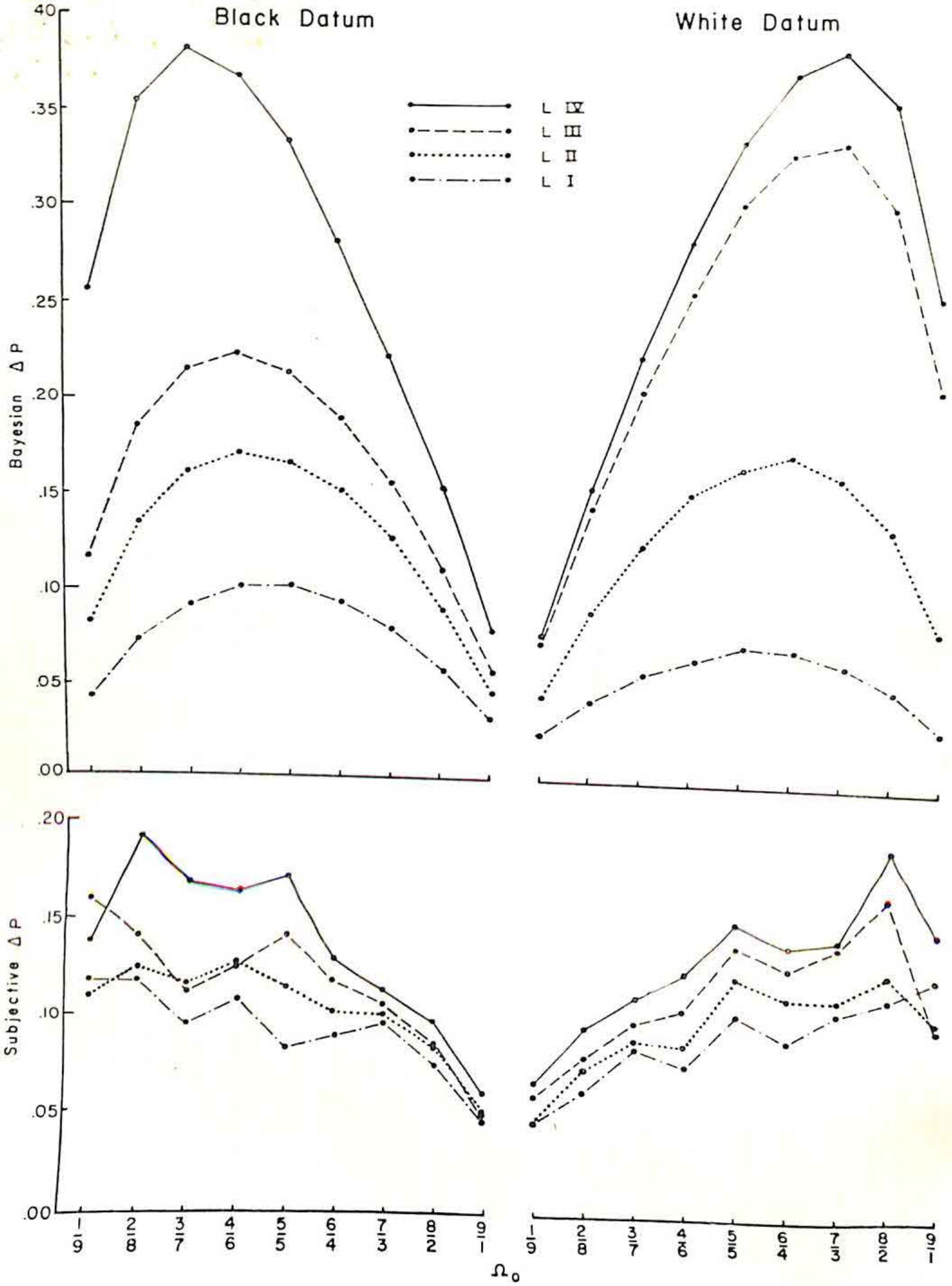


FIG. 2. A comparison of Bayesian and mean subjective ΔP as a function of Ω_0 and L. (Asymmetry of Bayesian ΔP with respect to a black vs. a white datum necessitated the left-right separation of the graphs. Since the datum that occurred was determined by the roll of a die, the size of the data base varies for different points on the graphs.)

REFERENCES

EDWARDS, W., LINDMAN, H., & PHILLIPS, L. D. Emerging technologies for making decisions. In, *New directions in psychology.* Vol. 2. New York: Holt, Rinehart, and Winston, 1965.

EDWARDS, W., LINDMAN, H., & SAVAGE, L. J. Bayesian statistical inference for psychological research. *Psychol. Rev.*, 1963, **70,** 193–242.

EDWARDS, W., & PHILLIPS, L. D. Man as transducer for probabilities in Bayesian command and control systems. In M. W. Shelly & G. L. Bryan (Eds.), *Human judgements and optimality.* New York: Wiley, 1964. Pp. 360–404.

PETERSON, C. R., SCHNEIDER, R. J., & MILLER, A. J. Sample size and the revision of subjective probabilities. *J. exp. Psychol.*, 1965, **69,** 522–527.

PETERSON, C. R., & ULEHLA, Z. J. Uncertainty, inference difficulty, and probability learning. *J. exp. Psychol.*, 1964, **67,** 523–530.

(Received April 20, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 1, 122–129

FALSE RECOGNITION PRODUCED BY IMPLICIT VERBAL RESPONSES[1]

BENTON J. UNDERWOOD

Northwestern University

200 words were read to 100 Ss at a 10-sec. rate. For each word S decided whether it had or had not been read earlier. Critical stimulus words were inserted in the list, these words being presumed to elicit specified implicit responses. Later in the list the assumed implicit response words were presented. For these latter words for 3 of 5 classes of words, false recognition was much higher than for control words. The greater the prior frequency of elicitation of the implicit response the greater the likelihood of false recognition.

In the development of conceptual schemes to incorporate verbal-learning phenomena, implicit responses are being given more and more prominent roles. Two examples may be cited. The studies of mediation in transfer evolved from two basic assumptions, namely, that implicit responses to verbal stimuli occur and that these implicit responses will serve as mediators. When conceptual similarity within a paired-associate list is found to produce interference it seems appropriate to account for this in part by assuming that the instances of a category (e.g., DOG, COW, HORSE) may all elicit the same implicit response (ANIMAL), which, in this situation, is a nondifferentiating response.

A distinction must be made between two kinds of implicit responses made to a verbal unit. There is first the response made to the unit itself as the act of perceiving it. This implicit response has been called the *representational response* by Bousfield, Whitmarsh, and Danick (1958), and will be abbreviated here as RR. The second kind of implicit response, the one with theoretical relevance, is produced by the stimulus properties of the RR. This implicit response may be another word which is associated with the actual word presented and will be called the *implicit associative response* (IAR). The particular IAR to a given word is often assumed to be the most frequent associate produced to the word in word-association procedures. Or, as in a mediating transfer situation, the IAR is assumed to be the word which *S* had learned as a response to a given stimulus in a previous list. That other nonverbal implicit responses (such as affective responses) may occur to the RR cannot be denied, but for the present purposes, need not be considered.

It must be clear that IAR, in most theoretical formulations, is conceived of as actually occurring. This is to say, it is not a hypothetical construct. It is hypothetical only in the sense that it is assumed to occur in a particular situation where it cannot be observed directly, and this assumption is made because it *has* been observed to occur overtly with a certain frequency in other situations (e.g., word-association procedures). The validity of the assumption is tested by experimental procedures wherein a given phenomenon is predicted *if* specific IARs are occurring.

---

[1] This work was done under Contract Nonr-1228 (15), Project NR 154-057, between Northwestern University and the Office of Naval Research.

The present study fits into this framework.

The rationale of the experiment may be understood by examining the task presented *S*s. A total of 200 words was read to *S* at a 10-sec. rate. As each word was read, *S* made a decision as to whether or not the word had been read to him earlier in the list. Thus, the general procedure is that devised by Shepard and Teghtsoonian (1961) in their study of memory for 3-digit numbers. In the present list of 200 words were critical words assumed to elicit particular IARs. At a later point in the list, therefore, the words assumed to be IARs to earlier RRs were in fact presented to *S*. If *S* responded by saying that these words had been read earlier in the list we would conclude that the IARs had occurred to the critical words at the time these critical words were read. For example, if a critical stimulus word is UP, and if later the assumed IAR (DOWN) is presented and *S* indicates that DOWN had occurred earlier in the list, we assume that DOWN had occurred as an IAR to the RR to the word UP, and that *S* subsequently confused an IAR with an RR.

METHOD

Words.—Four different types of words must be distinguished.

Critical Stimulus Words (CS Words): These words were assumed to elicit particular IARs.

Experimental Words (E Words): These were the words representing the assumed IARs to the CS Words.

Control Words (C Words): These words were used as controls for the E Words and were assumed not to have been preceded in the list by words for which they were IARs.

Filler Words (F Words): These words were presumed to be neutral with regard to the E Words and were used to build up a specific repetition frequency.

There were 20 E Words (hence, 20 C Words) as may be seen in Table 1. There were five different classes of CS Words, labeled A1, A3, CV, SO, and SI in Table 1. Each of these classes must be explained. Classes A1 and A3 consisted of words for which the E Word (assumed IAR) was an antonym of the CS Word. Thus, TOP is assumed to be the IAR to BOTTOM, TAKE to GIVE, and so on. In forming A1 and A3, 16 pairs of antonyms were divided randomly into four subgroups of four pairs each. Two of these subgroups were then assigned randomly to A1 and A3 as CS Words and E Words, the second member of the pairs of the other two subgroups as C Words. However, for these C Words, no CS Word appeared earlier in the list. Thus, when *S* was confronted with the C Word DOWN, this had not been preceded by UP. On the other hand, when *S* was confronted with the E Word TOP, it had been preceded by the CS Word, BOTTOM. The rationale of the study predicts that *S* will be more likely to say that TOP rather than DOWN had occurred earlier in the list.

The CS Words for the A1 class were presented only once; those for the A3 class were each presented three times prior to the appearance of the E Word. A comparison between these two classes will provide evidence on the role of frequency of IARs.

The third class of CS Words in Table 1 is CV, an abbreviation for converging associations. Each of the CS Words on a line is known to elicit the E Word with appreciable frequency in word-association norms. Thus, BREAD is a frequent response to both BUTTER and CRUMB, SLEEP a frequent response to both BED and DREAM, and so on. In selecting words for the first three classes in Table 1, major use was made of the Connecticut word-association norms as given by Bousfield, Cohen, Whitmarsh, and Kincaid (1961) with minor use of the Minnesota norms (Russell & Jenkins, 1954).

The fourth class of CS Words in Table 1 is labeled SO (superordinates). Each of the CS Words on a given line is a specific instance of the category which becomes the E Word. These Words were taken from Cohen, Bousfield, and Whitmarsh (1957). In actual fact these investigators asked *S*s to give specific instances to the category name. However, we are assuming that the instances will in turn elicit the category name as an IAR with appreciable frequency. Some justification for this assumption is found in word-association norms. For example, the Connecticut word-association norms show that 50% of the *S*s responded to the stimulus word CANARY with BIRD.

TABLE 1

Critical Stimulus Words, Experimental Words, and Control Words

Class	Critical Stimulus Words	E Word	Position	C Word
A1	BOTTOM	TOP	113	DOWN
	GIVE	TAKE	135	GOOD
	DAY	NIGHT	170	LOW
	MAN	WOMAN	188	RICH
A3	ROUGH	SMOOTH	154	WEAK
	FALSE	TRUE	162	DIRTY
	HARD	SOFT	178	SHORT
	SLOW	FAST	192	GIRL
CV	BUTTER, CRUMB	BREAD	129	BRIDGE
	BED, DREAM	SLEEP	155	SMILE
	SUGAR, BITTER, CANDY	SWEET	147	SALT
	ANIMAL, CAT, BARK	DOG	182	HORSE
	DARK, HEAVY, LAMP, MATCH	LIGHT	175	LEG
	WARM, CHILL, FREEZE, FRIGID, HOT, ICE	COLD	196	CLOUD
SO	MAPLE, OAK, ELM, BIRCH	TREE	123	FISH
	COTTON, WOOL, SILK, RAYON	CLOTH	158	FRUIT
	ROBIN, SPARROW, BLUEJAY, CANARY	BIRD	189	FLOWER
SI	BARREL, DOUGHNUT, DOME, GLOBE, SPOOL	ROUND	146	SHARP
	ATOM, CABIN, GERM, GNAT, VILLAGE	SMALL	165	FAT
	BANDAGE, CHALK, MILK, RICE, SNOW	WHITE	179	RED

The final class in Table 1 is SI (sense impressions). Taken from Underwood and Richardson (1956), each word on a line is known to elicit the E Word with appreciable frequency when the associations are limited to sense impressions by instructions and training.

The CS Words on a given line are not completely without associates to E Words on other lines. There are two fairly obvious instances. The CS Word DOUGHNUT in the SI class also elicits the word SWEET with appreciable frequency, SWEET being an E Word in the CV class. So also SNOW leads to COLD. Such "contamination" would, according to the notions in the introduction, tend to increase the number of false recognitions of the appropriate E Words.

The column labeled "Position" in Table 1 represents the position in the series of 200 words occupied by each E Word. The position of a given C Word was always two positions away from the position of its corresponding E Word. Thus, the C Word DOWN occurred at Position 111. Taken as a whole, the C Words occurred after the E Words half the time and before the E Words half the time.

The CS Words were scattered throughout the positions with the last occurring at Position 151. However, it may be of value to indicate the number of positions between the last CS Words and the E Words of the classes. For A1, the range for the four words was from 86 to 104 positions, with a mean of 96. For A3, the range was 36–89, with a mean of 53; for CV, a range of 30–60, with a mean of 42; for SO, from 27–68, with a mean of 50, and finally, for SI, 27–73, with a mean of 53.

There were 47 F Words. Of these 42 occurred twice, 4 occurred three times, and 1 occurred only once. Therefore, the 200 positions are accounted for as follows: CS Words, 63; E Words, 20; C Words, 20; and F Words, 97. The actual repetition was arranged so that in the first quarter (first 50 words) 5 words had occurred earlier, with 19, 16, and 18 having occurred earlier in each of the remaining three quarters.

Procedure and Ss.—The instructions and the words were presented by a magnetic tape recorder. The *S*s were given a single sheet of paper on which 200 numbered blanks occurred. The instructions required *S* to record a plus if he believed the word had occurred earlier in the list, a minus if he believed it had not. The instructions indicated that if *S* was in doubt a decision had to be made; a plus or minus had to be recorded for each blank. The words were read at a 10-sec. rate, each being spoken twice in immediate succession. The trial number was indicated after every tenth word to avoid any confusion as to the particular number of the word at the moment.

A total of 107 *S*s records was completed in four group sessions. All *S*s were college students taking introductory psychology courses at Northwestern. Seven of the records were eliminated on a random basis, so the data to be presented are based on 100 *S*s.

RESULTS

Table 2 shows the number of plusses (hence, percentage of *S*s) recorded for each E Word and its corresponding C Word. Also shown are the subtotals for each class of words. Looking at these subtotals it may be quickly noted that for Classes A1 and SI, expectations are not supported but that for Classes A3, CV, and SO, frequency of plusses for the E Words is much greater than for the C Words. However, we may first make an overall statistical evaluation without regard to classes of words. For each *S* the total number of plusses made to the 20 E Words was determined and also the total number made to the 20 C Words. The E Words produced a mean of 4.43 plusses, the C Words a mean of 2.53. The difference (1.90 ± .23) gives a *t* of 8.26. It must be concluded that *S*s responded more frequently to the E Words than to the C Words.

A statistical analysis of differences in frequency for E and C Words by classes of items is difficult by conventional methods since many *S*s did not respond with a plus to any of the words within a class. Even in the above analysis there were 8 *S*s with a zero entry for the 20 E Words and 24 *S*s with a zero entry for the C Words. However, certain simple computations allow some fairly firm conclusions about the statistical significance of the responses by classes of words, a matter to which we will turn shortly.

We may ask about the frequency of plusses for all words except E Words when they first occurred. This determination across the entire series of words provides a measure of the "false-alarm" or "false-positive" rate. The 200 words were divided into eighths and the number of false alarms determined for each of these eight sections. The number of words which first occurred (hence, could result in false alarms) in each successive eighth was 23, 22, 16, 15, 18, 9, 10, and 9. The mean numbers of plusses (false alarms) per word in each eighth are plotted as filled dots in Fig. 1 with a straight line drawn through them. It is clear that there is an increase in false alarms as the number of prior words increases; in the eighth

TABLE 2

NUMBER OF PLUSSES GIVEN TO EACH E AND C WORD BY 100 *S*s WITH SUBTOTALS FOR EACH CLASS OF WORDS

Class	E Word	No. of Plusses	C Word	No. of Plusses
A1	TOP	9	DOWN	20
	TAKE	11	GOOD	5
	NIGHT	24	LOW	27
	WOMAN	5	RICH	10
	Total	49		62
A3	SMOOTH	28	WEAK	9
	TRUE	12	DIRTY	9
	SOFT	37	SHORT	25
	FAST	49	GIRL	9
	Total	126		52
CV	BREAD	21	BRIDGE	28
	SLEEP	23	SMILE	10
	SWEET	24	SALT	11
	DOG	20	HORSE	7
	LIGHT	42	LEG	13
	COLD	39	CLOUD	15
	Total	169		84
SO	TREE	19	FISH	9
	CLOTH	18	FRUIT	11
	BIRD	38	FLOWER	11
	Total	75		31
SI	ROUND	7	SHARP	9
	SMALL	14	FAT	7
	WHITE	3	RED	8
	Total	24		24

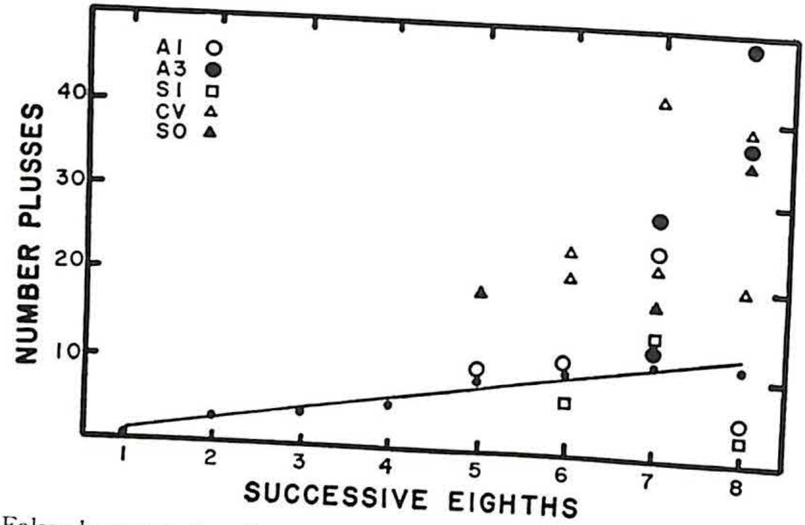


FIG. 1. False-alarm rate (number of plusses) within each successive block of 25 positions across the 200 words and number of plusses for each of the 20 E Words.

section the 100 *S*s averaged 11.2 false alarms for the nine words which first occurred in that section; in the first section the corresponding value was 0.5.

The plusses or false alarms for each E Word are also plotted in Fig. 1, above the eighth in which it occurred. The statistical question is whether or not the E Words of various classes can be considered to come from the population of words represented by the false-alarm rate of all non-E Words when they first occurred. Clearly, Classes A1 and SI could come from this population. It can be shown, however, that it is highly improbable that the E Words in each of the other three classes could have come from this population of false alarms for control words. The sum of the plusses for the three E Words in the SO class is 75; how probable is it that this sum would be equaled by any three control words used to derive the false-alarm rate in Fig. 1? To determine this the number of plusses for the 46 words used to derive the last half of the false-alarm curve were each listed. From this listing 100 samples of three each were drawn (using a table of random numbers) and the sum of the plusses for each sample determined. In drawing the samples, replacement was immediate; that is, the same value (representing the number of plusses for a given word) could have been drawn so as to constitute all three entries in a sample. The mean of this distribution of 100 sums was 30.71, with a standard deviation of 12.04. The maximum summed value obtained for any one sample was 71. Since the E Words in the SO class produced a sum of 75, we conclude that it is quite unlikely that these three E Words could come from the control population. This is to say, therefore, that the frequency of false alarms given to these E Words must have been influenced by the earlier presence of the CS Words. Or, to say this another way; *S* apparently confused the RR of the E Word with the IAR made earlier to the CS Word.

The above procedure could be repeated for the A3 and CV classes and the same conclusion would be reached. However, this is quite unnecessary. The maximum frequency of plusses for any word used in obtaining the control false-alarm rate was 28 (for BRIDGE, a C Word). Multiplying this value by 4 (112) does not yield a frequency as high (126) as the frequency produced by the four E Words in the A3 class. Multiplying 28 by 6 does not result in a frequency as high as that produced by the six E Words in the CV class.

In summary, therefore, the evidence indicates that RRs were confused with IARs for three classes of words. Among the 13 E Word-C Word comparisons in these three classes there is only one instance in which the frequency of the plusses for the C Word was higher than for the E Word (Class CV, BRIDGE vs. BREAD). The maximum frequency of false alarm for any E Word was 49; 49% of the *S*s indicated that the word FAST had occurred earlier in the list when in fact it had not. It is of some interest to compare this value with that of the lowest frequency of detection of a repeated word. The F Word HOME appeared in Position 13 and again in Position 118 at which time only 51% of the *S*s indicated it had occurred earlier. Overall, however, the correct detection of repeated words was 85.0%. It is evident, therefore, that *S* is far more likely to recognize a repeated word than to give a false alarm to a nonrepeated word even when we attempt to confuse him.

Other relationships.—The *S*s showed appreciable consistency in their tendency to report that a word had occurred earlier when in fact it had not. The product-moment correlation across the 100 *S*s for number of plusses for the 20 E Words and for the 20 C Words was .62. A total of 102 of the 200 positions was held by CS Words and by the first occurrence of F Words. The correlation between the number of plusses made to these 102 words and the number made to the E Words was .60, and to the C Words, .67. On the other hand, the ability to detect a repetition is not related to the tendency to indicate that a word had occurred earlier when it had not. There were 58 positions held by repeated words. The correlation across the 100 *S*s between the number of plusses assigned these 58 positions and the number of plusses assigned the 102 positions when words first occurred was −.02. A scatter plot gives no indication of curvilinearity.

DISCUSSION

Three of the classes of E Words produced results in conformance with the notion that RRs may be confused with IARs that occurred earlier; two classes did not. The fact that words in the A1 class did not produce false alarms beyond the control rate while those in A3 did, indicates that frequency of IARs is a critical variable. It would seem quite reasonable to presume that if an IAR occurs three times it has more of the properties of an RR than if it occurs only once. Indeed, it is precisely this fact which Deese (1959) has demonstrated by a sharp relationship between the appearance of an intrusion (a word which had not appeared in the list) in recall and the associative strength of the intruding word to the words actually in the list. However, in the case of the A1 words vs. the A3 words in the present study, another factor might be involved. As noted in the procedure section, on the average the number of words occurring between the CS Words and the E Words was greater for A1 than for A3. Perhaps the IARs to the CS Words in A1 are, therefore, forgotten or interfered with more than those produced by the CS Words in A3. While this possibility cannot be ruled out entirely, there is evidence against it. In the A3 class a total of 89 words occurred between the last presentation of the CS Word HARD and its E Word SOFT, yet 37% of the *S*s indicated they had heard the word earlier. For two of the cases in the A1 class, 86 and 89 words occurred between the CS Word and the E Word, yet the E Words produced only 11% and 5% plusses. Furthermore, an examination of the results for the CV class indicates that those cases where four and five CS words occurred (and were presumed to elicit a common associate) produced more false alarms than did the cases in the CV class where only two and three CS words were used, and this occurred when the number of words between the last CS and the E Word was essentially the same for all cases. Thus, the evidence points strongly to frequency of IAR as a variable in producing confusion between IARs and RRs.

The SI class of words gave completely negatively results; frequency of false alarms to the E Words in this class was no greater than for the C Words. As will be remembered, a given set of CS words in the SI class will, under appropriate instructions and training, elicit the common response which was here used as the E Word. It is possible that within the context of the present series of words these sense-impression responses occur as IARs with low frequency. However, in the Connecticut word-association norms (Bousfield et al., 1961) four of the five CS Words presumed to elicit the E Word ROUND do in fact elicit ROUND with some frequency (10%, 9%, 9%, and 28%). Nevertheless, these frequencies are much lower than the frequencies in the other classes. We may tentatively conclude that the negative results in the SI class are due to the failure of the CS words to elicit the common IARs with appreciable frequency.

The logic of the experiment was that the RR to the word of the moment would lead *S* to say that the word had occurred earlier if this RR was the same as an IAR produced by a word or words earlier in the list. This scheme might be reversed to say that the IAR to the word of the moment, if the same as an RR to an earlier word, would lead *S* to say that the present word had occurred earlier. Thus, if the word of the moment was DOWN, and if it elicits the IAR UP, and if UP occurred earlier in the list, *S* might believe that DOWN had occurred earlier. Such a mechanism might indeed handle the results for A3, for antonyms are bidirectionally associated. And, such a mechanism might account for the failure for the SI items, since the E Words do not elicit the CS Words with appreciable frequency in available norms. However, it cannot account for the results for the CV and SO classes. The positive effects produced in these classes must be due to the elicitation of the same IAR by two or more different CS Words, otherwise positive results would have been obtained in the A1 class. The IAR of the word of the moment cannot be confused with two or more different RRs of previous words such as would be necessary by the alternative scheme. Of course, it is possible that both mechanisms may be involved but until this is clearly demonstrated we will conclude that the original scheme presented is most appropriate; the RR to the word of the moment is confused with the common IAR produced by two or more previous words. This is essentially the same conclusion reached by Mink (1963) as a result of his studies on semantic generalization. And it should be noted that the present study could be classified as a study of secondary or semantic generalization.

Figure 1 showed that the frequency of false alarms increased throughout. Although this increase may be due to the same mechanisms presumed to operate for the E Words, another factor may be in part responsible. It is possible that the greater the number of words *S* heard the more likely he would believe that a repetition would occur. Thus, if in doubt late in the list he would be more likely to signify that a repetition had occurred than if in doubt early in the list. This could be true in spite of the fact that the objective repetition was held constant over the last three quarters of the list. Nevertheless, certain of the C Words produced false alarms with such high frequency that an examination of possible causes was made. Some may have been produced by high formal similarity of RRs. For example, LOW produced 27 false alarms. The word LAW occurred as an F Word earlier in the list. The high frequency for SHORT may have been due to the appearance of SHARP earlier in the list. The C Word which produced the highest number of false alarms was BRIDGE. No previous word has high formal similarity and, when BRIDGE is thought of as a structure spanning water, no previous word would seem to elicit it as an IAR. If these *S*s are avid players and followers of the game of bridge, such words as MASTER and MAJOR (which occurred three times) may have elicited BRIDGE as an IAR. In any event, it is apparent that we were unsuccessful in the attempt to establish homogeneity in false-alarm rates to control words.

REFERENCES

BOUSFIELD, W. A., COHEN, B. H., WHITMARSH, G. A., & KINCAID, W. D. The Connecticut free association norms. Technical Report No. 35, 1961, University of Connecticut, Contract Nonr-631 (00), Office of Naval Research.

BOUSFIELD, W. A., WHITMARSH, G. A., & DANICK, J. J. Partial response identities in verbal generalization. *Psychol. Rep.*, 1958, **4**, 703–713.

COHEN, B. H., BOUSFIELD, W. A., & WHITMARSH, G. A. Cultural norms for verbal items in 43 categories. Technical Report No. 22, 1957, University of Connecticut, Contract Nonr-631 (00), Office of Naval Research.

DEESE, J. On the prediction of occurrence of particular verbal intrusions in immediate recall. *J. exp. Psychol.*, 1959, **58**, 17–22.

MINK, W. D. Semantic generalization as related to word association. *Psychol. Rep.*, 1963, **12**, 59–67.

RUSSELL, W. A., & JENKINS, J. J. The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Word-Association Test. Technical Report, No. 11, 1954, University of Minnesota, Contract N8 onr-66216, Office of Naval Research.

SHEPARD, R. N., & TEGHTSOONIAN, M. Retention of information under conditions approaching a steady state. *J. exp. Psychol.*, 1961, **62**, 302–309.

UNDERWOOD, B. J., & RICHARDSON, J. Some verbal materials for the study of concept formation. *Psychol. Bull.*, 1956, **53**, 84–95.

(Received April 21, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 1, 130–132

SUPPLEMENTARY REPORT

PRIMACY AND RECENCY EFFECTS IN SERIAL-POSITION CURVES OF IMMEDIATE RECALL[1]

JOHN C. JAHNKE
Miami University

Each of 48 college students were read lists of 6, 10, and 15 English words in counterbalanced order. Half the Ss were given instructions for the immediate serial recall of the lists; the remaining, immediate free recall. Mean percentage correct recall was found to be higher for conditions of free than serial recall and for shorter than longer lists. Serial-position curves showed that recency effects, relative to primacy, were stronger for free than serial recall and for longer than shorter lists. Frequency of recall appeared to be more closely related to order of emission of an item during recall than to its order during presentation.

Deese (1957) has presented data indicating that the form of the serial-position curve of immediate free recall is different from that of immediate serial recall. The first purpose of the present study is to verify this finding. Because other studies have examined serial-position curves of immediate free recall only with relatively long lists of items (e.g., Deese, 1957; Deese & Kaufman, 1957; Murdock, 1962) and of immediate serial recall only with relatively short lists (e.g., Jahnke, 1963), the present study also examined recall with both long and short lists.

Method.—The *S*s were 48 Miami University undergraduate students of both sexes. None of the *S*s had served previously in any experiment on immediate memory.

The *S*s were instructed for immediate recall of lists of English words. Half the *S*s were told to listen to the words and recite as many of the words as they could in the order they heard them (serial recall). The remaining *S*s were told to listen to the words and recite as many of them as they could (free recall). Subject to the condition that the ratio of men to women be approximately the same in each recall condition, assignment of *S*s to the recall conditions was made in the abba sequence.

List lengths used here were 6, 10, and 15 words. In both recall conditions, *S*s received a total of 12 different lists, 4 at each of the three lengths. One trial was given on any one list. Twelve different orders of presentation of the lists were prepared such that each length of list appeared once in each block of three trials. Any order of presentation in a three-trial block was never repeated over the four blocks for any one *S*. Each list length appeared approximately equally often at each stage of practice. The *S*s were informed before each trial how many words would be in the list they would next hear.

The words of each list were chosen by drawing without replacement from the pool of those words with frequency 45–49 per million (Thorndike & Lorge, 1944) in order to keep the average frequencies of the lists of different lengths essentially constant. After a thorough shuffling, the lists at each length were used one after another as required on successive trials. The words were read at a 1/sec rate without inflection or emphasis by a female *E*, who also recorded *S*'s responses during the 30-sec. recall interval which followed each list.

Results.—Mean percentage correct recall, given in order of increasing list length, was 81.94, 52.60, and 38.12 for the free-recall condition and 76.56, 45.73, and 29.79 for the serial-recall condition. A mixed-factorial analysis of variance of these data showed that recall was better for free than serial recall and for shorter than longer lists, $F\ (1, 46) = 7.37$ and $F\ (2, 92) = 626.63$, respectively, $p < .01$. The interaction effect was not significant ($F < 1$). For these and all subsequent analyses, a word recalled was scored correct if it was from the list to be recalled, even if the word was not given in its correct serial position.

Serial-position curves for each experimental condition are presented in Fig. 1 in terms of

---

[1] This study was supported in part by Contract AF 33(657)-10456 with the Aerospace Medical Laboratories, Engineering Psychology Branch, Wright-Patterson Air Force Base, Ohio, and in part by a summer research appointment to the writer from Miami University.

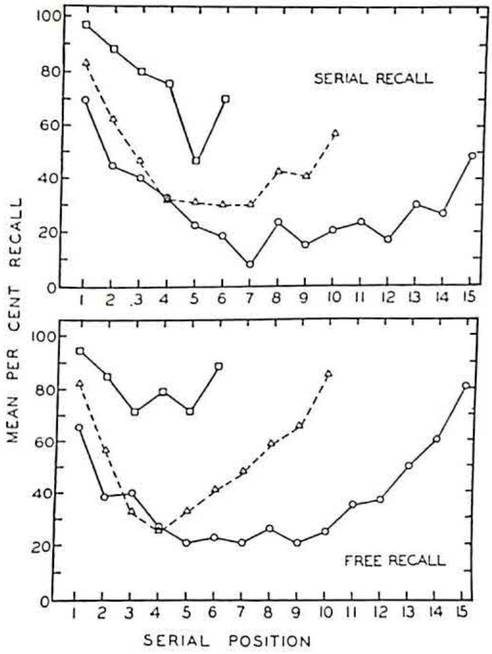


FIG. 1. Serial-position curves for each experimental condition.

performance averaged over the four trials at each length of list. Inspection of Fig. 1 suggests that the relative strengths of primacy and recency effects are different for the two recall conditions and also, possibly, for the lists of different lengths. To test these relations, the number of times the last two items were given correctly (a measure of recency) was subtracted from the number of times the first two items were given correctly (a measure of primacy) for each *S* for each length of list. These difference scores were subjected to a mixed-factorial analysis of variance, which showed that the main effects of recall condition and list length were significant, $F\ (1, 46) = 20.79$ and $F\ (2, 92) = 7.80$, respectively, $p < .01$. The interaction effect was not significant ($F < 1$). These results show that recency effects, relative to primacy, are stronger for free than serial recall and for longer than shorter lists. The nonsignificant interaction shows that the trend toward relatively stronger recency effects with increasing list length was much the same for both recall conditions of the present study.

Studies of immediate serial recall and immediate free recall of lists of unrelated words (Bousfield, Whitmarsh, & Esterson, 1958; Deese, 1957; Deese & Kaufman, 1957) have shown that there is a high positive correlation between the frequency with which an item is recalled (FR) and the order in which an item is emitted during recall (OR). Such correlations were computed separately for each of the six experimental conditions of the present study, as were also the correlations between FR and the order of an item during presentation (OP) and between OP and OR. Table 1 presents these rank-order correlation coefficients. Table 1 shows that FR and OR are, in general, strongly related. The one nonsignificant correlation between these two variables, that for the free recall of the 6-item lists, may be explained in that items tended to be ordered in recall as they were in presentation ($\rho = .943$, $p < .05$), but that the last item was very frequently recalled. With but one other exception, not only were the correlations between FR and OR very high, but all other correlations were low.

The OP-OR correlations seen in Table 1 suggest that *S*s order their recall differently in serial and free recall. Inspection of the raw data showed that when recall was serial, *S*s tended to recall in the order given, particularly on the first few trials. With longer lists, those items likely to be recalled after the first few were those from the end of the list. When recall was free, *S*s tended to recall items in the order given only when lists were short. With longer lists, recall was usually of last items first, followed frequently by first items.

Finally, since *S*s in the serial-recall condi-

TABLE 1

RANK-ORDER CORRELATION COEFFICIENTS BETWEEN FREQUENCY OF RECALL (FR) AND ORDER OF AN ITEM DURING PRESENTATION (OP) AND RECALL (OR)

Variables	Free Recall			Serial Recall		
	6 Items	10 Items	15 Items	6 Items	10 Items	15 Items
FR-OR	.529	.976**	.982**	1.000**	.985**	.989**
FR-OP	.329	−.355	−.204	.943*	.367	.239
OP-OR	.943*	−.173	−.296	.943*	.358	.296

* $p < .05$.
** $p < .01$.

tions were requested to recall the items in order of presentation, they might be expected not to verbalize remembered words when they were unsure of the words' positions. Consequently, both fewer correct responses and more omissions would be expected for the serial-recall conditions. A significantly smaller percentage of correct responses was noted above for the serial-recall than the free-recall condition. The mean numbers of omissions were also significantly different, with differences in the expected direction, $t\ (46) = 2.618,\ p < .01$.

Discussion.—The present study confirms and extends earlier findings concerning the effect of instructions for immediate free and serial recall on the form of the serial-position curve. When instructions were given for serial recall, primacy effects were relatively stronger and recency effects relatively weaker than when instructions were given for free recall (cf., Deese, 1957; Deese & Kaufman, 1957; Jahnke, 1963; Murdock, 1962). For both recall conditions recency effects became stronger, relative to those of primacy, as length of list increased.

REFERENCES

BOUSFIELD, W. A., WHITMARSH, G. A., & ESTERSON, J. Serial position effects and the "Marbe effect" in the free recall of meaningful words. *J. gen. Psychol.*, 1958, **59**, 255–262.

DEESE, J. Serial organization in the recall of disconnected items. *Psychol. Rep.*, 1957, **3**, 577–582.

DEESE, J., & KAUFMAN, R. A. Serial effects in recall of unorganized and sequentially organized verbal material. *J. exp. Psychol.*, 1957, **54**, 180–187.

JAHNKE, J. C. Serial position effects in immediate serial recall. *J. verbal Learn. verbal Behav.*, 1963, **2**, 284–287.

MURDOCK, B. B., JR. The serial position effect of free recall. *J. exp. Psychol.*, 1962, **64**, 482–488.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, Bureau of Publications, 1944.

(Received March 30, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 1, 133–134

Beginning with the first issue of Volume 69 and continuing through the remainder of that volume and Volume 70 (1965), the titles and authors of accepted papers will be listed here following Supplementary Reports. It is being supported on an experimental basis by the APA Project on Scientific Information Exchange in Psychology, and at the end of the year, the outcome of this trial will be evaluated and consideration given the advisability of continuing the listing.

This listing plus those published in the preceding 1965 issues are a portion of the backlog of manuscripts accepted by this journal. Such listing will allow readers to become aware of research many months in advance of journal publication.

Manuscripts Accepted for Publication in the *Journal of Experimental Psychology*

Facilitation of Performance Through the Use of the Timing System: Stuart J. Dimond*: 57 Woodbridge Road, Knowle, Bristol 4, England.

Absolute Judgments of Duration: Lawrence E. Murphy*: Department of Psychology, University of Arizona, Tucson, Arizona 85721.

Errors and Error Correction in Choice-Response Tasks: Patrick Rabbitt*: Medical Research Council, Applied Psychology Research Unit, 15 Chaucer Road, Cambridge, England.

Time and Event Uncertainty in Unisensory Reaction Time: Donald Reynolds*: HumRRO, Division No. 1 (System Operations), George Washington University, 300 North Washington Street, Alexandria, Virginia 22314.

Effect of Noncontingent Partial Reinforcement on the Resistance to Extinction of a Runway Response: Milton A. Trapold* and David G. Doren: Department of Psychology, University of Minnesota, Minneapolis, Minnesota 55455.

Percentage Reinforcement and Reward-Magnitude Effects in a T Maze: Between and Within *S*s: Norman E. Spear* and William B. Pavlik: Department of Psychology, Rutgers University, New Brunswick, New Jersey 08903.

Predicting Discrimination Learning from Differential Conditioning with Amount of Reinforcement as a Variable: R. A. Champion* and L. R. Smith: Department of Psychology, University of Sydney, Sydney, N. S. W., Australia.

Transfer of Predifferentiation Training to Gradients of Generalization in Shape Recognition: Henry C. Ellis* and Robert L. Feuge: Department of Psychology, University of New Mexico, Albuquerque, New Mexico.

Determinants of the Effects of Vicarious Reinforcement: Albert R. Marston*: Department of Psychology, University of Wisconsin, Charter at Johnson, Madison, Wisconsin 53706.

What Happens to the Stimulus in Backward Masking?: John H. Thompson*: Department of Psychology, Gonzaga University, Spokane, Washington 99202.

Effects of Reinforcement History Upon Risk-Taking Behavior: Marshall G. Greenberg* and Bernard Weiner: Department of Psychology, University of Minnesota, Minneapolis, Minnesota 55455.

Stimulus Exposure Time in Paired-Associates Learning: Richard M. Good*: Division of Institutional Research, University of Maryland, College Park, Maryland 20742.

Effect of Stimulus Similarity on Retroactive Masking: Elizabeth Fehrer*: Department of Psychology, Brooklyn College of the City University of New York, Brooklyn, New York 11010.

Vibrotactile Adaptation and Recovery Measured by Two Methods: J. F. Hahn*: Psychological Laboratory, Gilmer Hall, University of Virginia, Charlottesville, Virginia.

Association by Contiguity: Role of Response Availability: Geoffrey Keppel*: Department of Psychology, University of California, Berkeley, California 94720.

Response Transfer as a Function of Verbal-Association Strength: Group Verbal Learning: Charles Clifton, Jr.*: Institute of Child Behavior and Development, University of Iowa, Iowa City, Iowa.

---

* Asterisk indicates author for whom address is supplied.

Leadership in Small Groups: A resolution of Discordance: Arnold Binder,* Burton R. Wolin, and Stanley J. Terebinski: Department of Industrial Engineering and Operations Research, New York University, University Heights, Bronx 53, New York.

Information versus Reward in Binary Choices: Amos Tversky and Ward Edwards*: Engineering Psychology Laboratory, Institute of Science and Technology, University of Michigan, P. O. Box 618, Ann Arbor, Michigan 48107.

Sequential Effects in Disjunctive Reaction Time: Implications for Decision Models: Judith A. Williams*: Department of Psychology, McGill University, Montreal, Canada.

Starting Position, Adaptation, and Visual Framework as Influencing the Perception of Verticality: Ricardo B. Morant* and Joel Aronoff: Department of Psychology, Brandeis University, Waltham 54, Massachusetts.

Response Force as a Function of Amount of Reinforcement: Vincent Di Lollo,* W. D. Ensminger, and J. M. Notterman: Department of Psychology, Indiana University, Bloomington, Indiana 47405.

Stimulus Meaning and Complexity as Factors in the Transfer-of-Stimulus Predifferentiation: Henry C. Ellis,* Douglas G. Muller, and Donald T. Tosti: Department of Psychology, University of New Mexico, Albuquerque, New Mexico.

Stimulus Meaningfulness and Unlearning in the A-B, A-C Transfer Paradigm: Joseph A. Bryk and Donald H. Kausler*: Department of Psychology, St. Louis University, 221 North Grand Boulevard, St. Louis 3, Missouri.

Forward and Backward Associations in Paired Associates: Bennet B. Murdock, Jr.*: Department of Psychology, University of Missouri, Columbia, Missouri.

Mediated Stimulus Equivalence as a Function of the Number of Converging Stimulus Items: Donald J. Foss* and James J. Jenkins: Center for Advanced Study in the Behavioral Sciences, 202 Junipero Serra Boulevard, Stanford, California 94305.

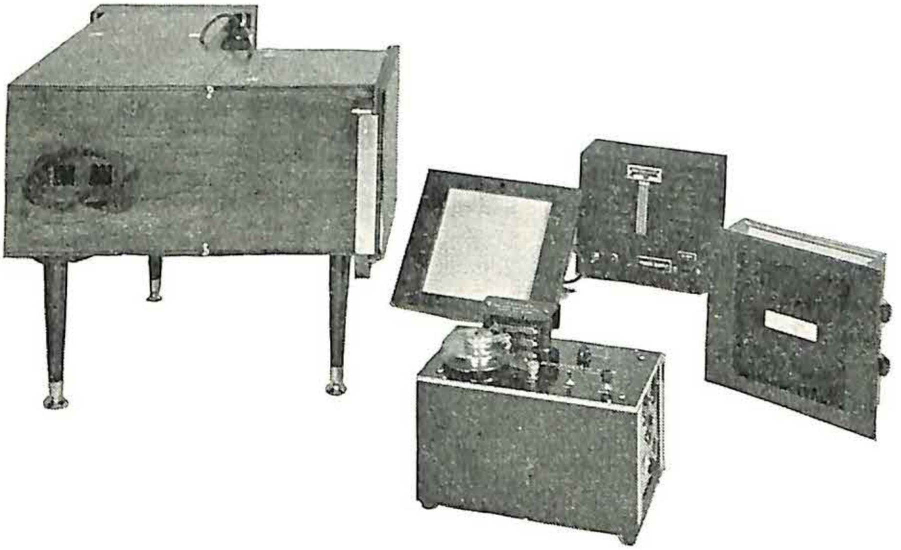
Acoustic Facilitation of Visual Detection: William H. Watkins* and Carl E. Feehrer: Hq ESD (ESRHD) Stop 36, LG Hanscom Field, Bedford, Massachusetts 01731.

First-List Retention and Time and Method of Recall: John P. Houston*: Department of Psychology, University of California, Los Angeles, California 90024.

Simple Reaction Time as a Function of the Relative Frequency of the Preparatory Interval: Theodore P. Zahn* and David Rosenthal: Laboratory of Psychology, National Institute of Mental Health, Bethesda 14, Maryland.

TACHISTOSCOPE

MODEL T-1C



A PROFESSIONAL* **QUALITY** INSTRUMENT FEATURING FAST **WHITE** LAMPS, LARGE EXPOSURE FIELDS, ACCURATE TIMER, VERSATILE ACCESSORIES FOR PRESENTING SINGLE CARDS, MULTIPLE CARDS, MATERIAL ON SPOOLED PAPER, TRANSLUCENT MATERIAL ILLUMINATED FROM BEHIND, ETC.

* P.S.—OUR LOW VOLTAGE SOLID STATE CIRCUITRY PERMITS SEPARATION OF COMPONENTS; NO HEAT, NO FANS, NO DISTRACTING NOISES.

COMPLETE AS SHOWN, $800.00 F.O.B. ARLINGTON

For Complete Data Write or Phone

Ralph Gerbrands Company

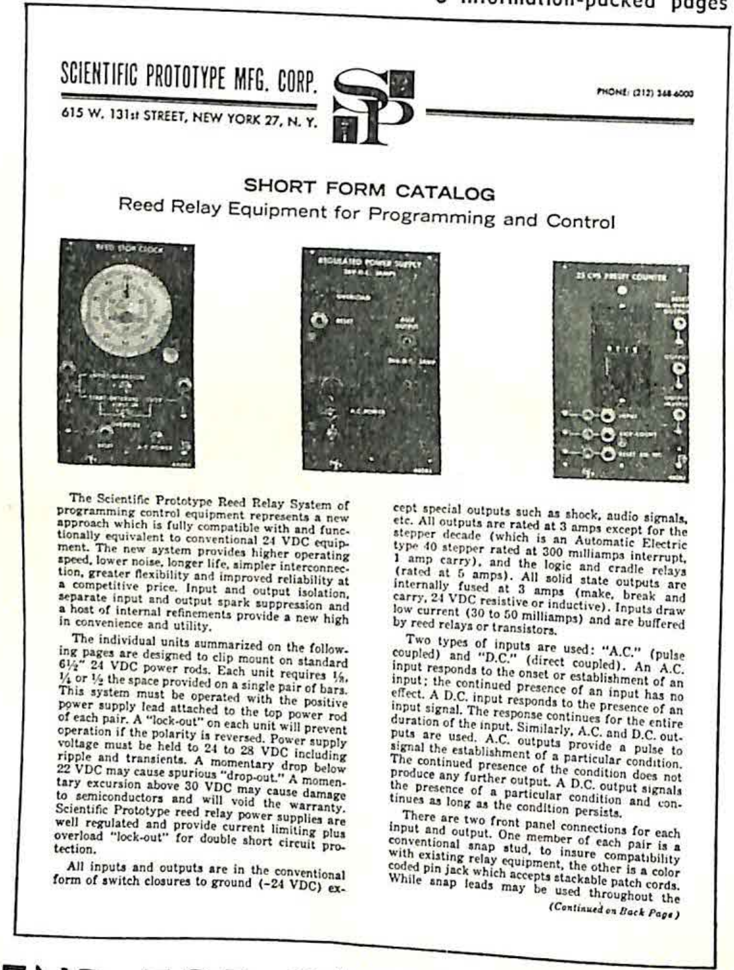
Scientific Instruments

8 Beck Road
Arlington, Mass. 02174

Tel. 648-6415
Area Code 617

NEW REED RELAY CONTROL EQUIPMENT

6 information-packed pages



SCIENTIFIC PROTOTYPE MFG. CORP.

615 W. 131st STREET, NEW YORK 27, N. Y.

PHONE: (212) 348-6000

SHORT FORM CATALOG

Reed Relay Equipment for Programming and Control

The Scientific Prototype Reed Relay System of programming control equipment represents a new approach which is fully compatible with and functionally equivalent to conventional 24 VDC equipment. The new system provides higher operating speed, lower noise, longer life, simpler interconnection, greater flexibility and improved reliability at a competitive price. Input and output isolation, separate input and output spark suppression and a host of internal refinements provide a new high in convenience and utility.

The individual units summarized on the following pages are designed to clip mount on standard 6½" 24 VDC power rods. Each unit requires ⅛, ¼ or ½ the space provided on a single pair of bars. This system must be operated with the positive power supply lead attached to the top power rod of each pair. A "lock-out" on each unit will prevent operation if the polarity is reversed. Power supply voltage must be held to 24 to 28 VDC including ripple and transients. A momentary drop below 22 VDC may cause spurious "drop-out." A momentary excursion above 30 VDC may cause damage to semiconductors and will void the warranty. Scientific Prototype reed relay power supplies are well regulated and provide current limiting plus overload "lock-out" for double short circuit protection.

All inputs and outputs are in the conventional form of switch closures to ground (–24 VDC) except special outputs such as shock, audio signals, etc. All outputs are rated at 3 amps except for the stepper decade (which is an Automatic Electric type 40 stepper rated at 300 milliamps interrupt, 1 amp carry), and the logic and cradle relays (rated at 5 amps). All solid state outputs are internally fused at 3 amps (make, break and carry, 24 VDC resistive or inductive). Inputs draw low current (30 to 50 milliamps) and are buffered by reed relays or transistors.

Two types of inputs are used: "A.C." (pulse coupled) and "D.C." (direct coupled). An A.C. input responds to the onset or establishment of an input; the continued presence of an input has no effect. A D.C. input responds to the presence of an input signal. The response continues for the entire duration of the input. Similarly, A.C. and D.C. outputs are used. A.C. outputs provide a pulse to signal the establishment of a particular condition. The continued presence of the condition does not produce any further output. A D.C. output signals the presence of a particular condition and continues as long as the condition persists.

There are two front panel connections for each input and output. One member of each pair is a conventional snap stud, to insure compatibility with existing relay equipment, the other is a color coded pin jack which accepts stackable patch cords. While snap leads may be used throughout the

(Continued on Back Page)

SEND FOR CATALOG TODAY

SCIENTIFIC PROTOTYPE MFG. CORP.

Behavior Analysis Apparatus . . . operant conditioning kits, rodent test cages, pellet dispensers, soundproof chambers, recorders, noise generators, shock sources, timers, switches, digital test equipment and special systems.

615 W. 131 Street, New York, N. Y. 10027
Telephones: (212) MO 6-4674 and 4789

Journal of
Experimental Psychology

VOL. 70, NO. 2 AUGUST 1965

TRANSFER OF TRAINING ACROSS TARGET SIZES[1]

INA McD. BILODEAU

Tulane University

The present paper deals with an instructional manipulation of target size or width of tracking path in an S-paced tracking task: E verbally defined the limits of correct responding, the limits varying from treatment to treatment. After extended training under one path width, S was tested with the same or a different path width. 2 experiments completed the 9 factorial combinations of the 3 values of path width used. Transfer of training was positive in all cases, but the amount of transfer declined as the difference between training- and test-target limits increased. Very regular generalizationlike functions were obtained when amount of transfer was plotted against magnitude of the shift in path width; direction of the shift was apparently irrelevant.

When perceptual-motor tasks can be scaled along some dimension (gear ratio, target speed, control pressure, etc.), transfer of training from one task to another should be a function of the difference between the tasks along the dimension (Osgood, 1949). Where positive transfer is expected, optimal performance is predicted when training and test tasks are the same, and lesser levels of performance, the greater the shift in treatment. Though there is a good deal of evidence that transfer of training is independent of target size, there is very strong evidence with simple positioning devices that the problem in these cases is to account for a behavioral insensitivity to target size rather than for an anomaly of transfer of training (Bilodeau, 1952a, 1952b, 1954, 1955, 1956). The present prediction that transfer of training will depend on shift in target size is based, therefore, on the expectation that the responses practiced during training can be made to differ from the responses required in a test.

In the present study target size was varied as width of the tracking pathway, or the latitude allowed in responses called correct. The task was to turn two control handles to drive a pin along a path; instructions defined the limits of correct responding. The study was designed to avoid two common difficulties in work with target size. The difficulties are: (*a*)

---

[1] The data on which this report is based were collected at Lackland Air Force Base, San Antonio, Texas.

The author wishes to acknowledge the assistance of Frank E. McClammy who ran most of the *S*s and who helped formulate the instructions.

failing to distinguish between scoring and performance by confounding scoring and target areas and (*b*) using tasks which do not penalize *S* if he fails to change his behavior with his target. The latter difficulty may explain why target size is irrelevant in transfer of training; the former makes it impossible to state whether the 100% transfer occurs despite differences in behavior from target to target, or because there are none (Bilodeau, 1952a, 1954). In our *S*-paced task, a difference in distance around the track necessitates some difference in performance; differences are encouraged by the fact that *S*'s score suffers if he responds within either narrower limits (he unnecessarily restricts adequate responses and loses speed) or wider limits (he must retrace to correct errors) than his target requires. That *S* must balance speed and accuracy for a high score implies also that transfer will be controlled more by amount than by direction of shift in target size.

A preliminary experiment with the present technique gave differences in training but not in a test after a half-dozen training trials. For caution, therefore, because of the investment of 9 days of practice per *S* and the novel manipulation of target size, the first of the following experiments tested for transfer with only the narrowest path width; the purpose was to determine if transfer would vary with shift in target size after extended training. The second experiment was then run to provide data under the two broader path widths to test the expectation that transfer would depend more on amount than direction of shift in target size.

METHOD

Subjects.—Permanent party airmen at Lackland Air Force Base served as *S*s, 48 in Exp. I, 66 in Exp. II. The *S*s were scheduled in pairs to run at about the same time of day throughout the experiment. Unsystematic testing sequences were used to assign the pairs to groups of 16 *S*s in Exp. I and 10 or 12 *S*s in Exp. II.

Apparatus.—The device used was the F unit of the Two-Hand Track Apparatus (Bilodeau & Bilodeau, 1958). The Two-Hand Track replaces the target of the Standard Two-Hand Coordination Test (Melton, 1947) with a path in the shape of a four-leaf clover and substitutes a tracking pin (.44 in. in diameter) for the target follower. The path is formed of 48 contacts embedded in a .5 × 19 in. plastic strip. The object is to drive the pin over the pathway, under instructions to trace the path as quickly as possible.

Procedure.—The procedure was designed to promote high levels of learning and performance before testing for transfer. Two *S*s in the same group alternated trials, one practicing while the other rested. The *S*s had five 1-min. trials a day at 80-sec. intertrial intervals for 9 days. For the first *S* of a pair, *E* moved the pin to the start of the path; thereafter *S* took over where his partner ended, except that *E* moved the pin back to the nearest point on the path if the partner finished with an error. Partners watched each other practice except on days introducing new treatments. The percentage of improvement of the pair as a team was reported before a day's practice began. The percentage compared the last trial of the preceding day with the last trial of the day before that; last vs. first trial of the preceding day was reported on the day after a change in treatment.

There were three treatments. Treatment P (precise), using the smallest target, was the standard treatment for which the apparatus was designed: *S* activated a stepping switch by centering the pin over successive contacts; the response tolerance in handle turning was 188° (±94° about the center of the contact). In treatments requiring less precision, the apparatus was turned off so that the stepping switch did not operate. The *S*s of Treatment M (moderate) met the target requirements so long as any part of the pin touched a contact, or, effectively the tolerance was about 865° in handle turning; *S* was to retrace to correct errors. Treatment F (free) required only that some part of the pin be kept in contact with some part of the track, so that the tolerance was approximately 1626°. Moreover, *S*s of Treatment F were not obliged to stop or retrace to correct an error—if the pin lost contact with the path, *S* was to bring the pin back as soon as possible to some point farther along in the path. In summary, the range of treatments was such that for a correct response in Treatment P an *S* had to position the pin at target center with accuracy within about a quarter turn of his crank handles, while in Treatment F, *S* could be inexact in hitting path center by as much as two and a quarter turns. As the task and instructions varied, *E*'s demonstration of correct and incorrect responding varied appropriately. Questions were asked to check *S*'s understanding. All three treatments stressed speed. Treatment F gave additional speed-up warnings on Training Day 1: to both *S*s after the second partner's Trial 2 and, individually, halfway through the second partner's Trial 3 and the first partner's Trial 4. The response measure was the distance traveled in number of circuits of the path per trial.

The *S*s received one of the three treatments for 6 days (training), and then had a 3-day test for transfer. The *S*s of Exp. I were tested under Treatment P after training under Treatment P, M, or F. Experiment II completed the factorial design for training and test combinations: about half the *S*s trained under one treatment were tested under Treatment M, the others under Treatment F. Table 1 provides the *n*'s with a summary of the experimental design.

TABLE 1

EXPERIMENTAL DESIGN

Exp.	Test	Training		
		P	M	F
I	P	16[a]	16	16
II	M	12	10	10
II	F	10	12	12

Note.—Entries give group *n*.
[a] $n = 15$ on Day 9, because one *S* was ill.

RESULTS AND DISCUSSION[2]

Mean number of circuits of the path (19 in. along path center) per trial is plotted for both training and test in Fig. 1, breaks in the curves separating the daily sessions. The data in Fig. 1 will be briefly considered before the major comparison of intertreatment transfer is presented.

Training.—As the two experiments yielded comparable trends, their training data were averaged for the curves in the left-hand portion of Fig. 1. For present purposes, the most important point is that the treatments were adequate in producing widely different training experiences: on all 30 trials the means are ranked in order of path width. But there is also a treatment effect on the trends; the function relating performance to trials is nearly linear in Treatment P, and negative acceleration becomes more marked as target size increases and scoring difficulty decreases from P to M to F. A comparison of training treatments, using the ranks of *S* totals for Day 6 of training, yielded $H(2) = 92$, $p < .01$ (Kruskal & Wallis, 1952).

Test.—The right-hand portion of Fig. 1 presents the performance of the nine groups in the transfer-test trials of Days 7–9. The symbol identifies test, the line, training treatment. For example, Group P-M combines solid lines (P training) with Xs (M testing), Group F-P combines dotted lines (F training) with closed circles (P testing), and Group M-F combines dashed lines (M training) with open circles (F testing). Several points can be made by inspection. (*a*) Transfer of training is positive in every instance (compare early trials of test and training curves using the same symbol). (*b*) Transfer is not complete (compare test trials of groups having the same symbol); for any test treatment, performance is best when *S* had that treatment in

---

[2] Tables of trial means and variances have been deposited with American Documentation Institute. Order Document No. 8385 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington, D. C. 20540. Remit in advance $1.25 for microfilm or $1.25 for photocopies and make checks payable to Chief, Photoduplication Service, Library of Congress.

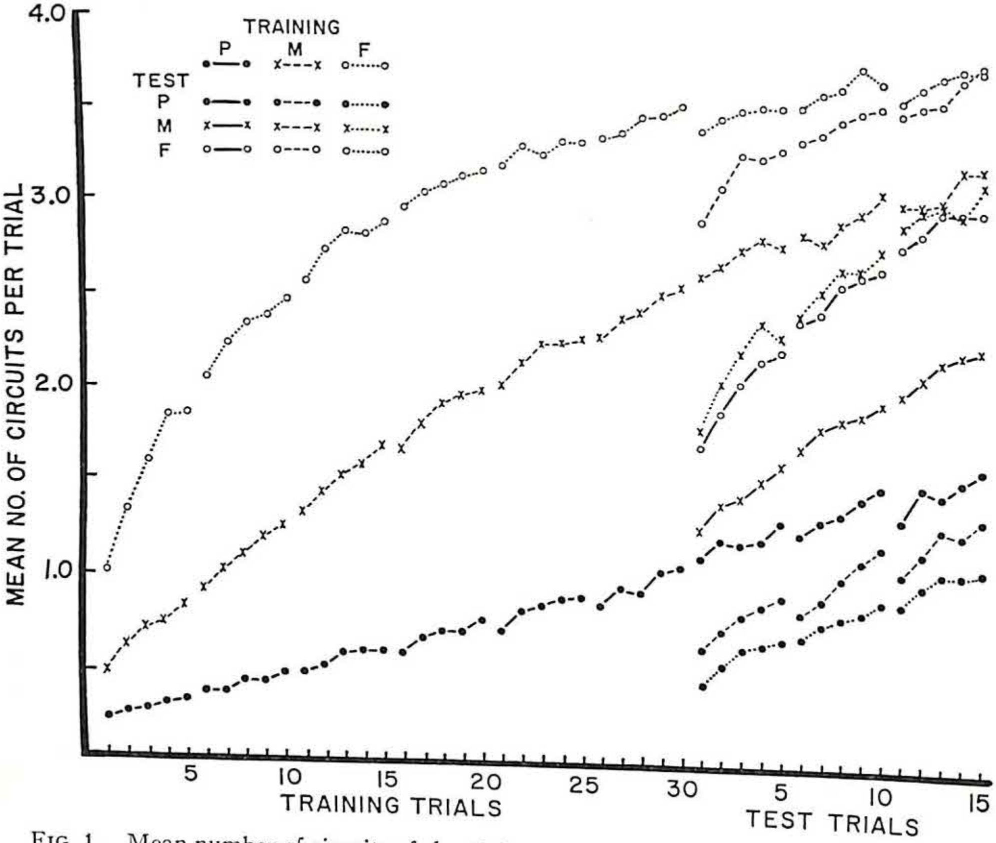


FIG. 1. Mean number of circuits of the 19-in. tracking path on 30 training trials for the three training treatments and on 15 test trials for the nine groups.

training. (*c*) Transfer to the extreme test treatments, P and F, decreases, the greater the change in path width from training to test. (*d*) The curves of groups shifted between M and F converge as the test progresses, shifts between P and either M or F produce less ready adjustment to the new task. (For each symbol, compare solid lines with the two other curves.) Even the early test trials suggest greater intertask transfer between M and F, in that F-trained *S*s are superior to P-trained on M, the intermediate treatment (Xs, dotted lines vs. Xs, solid lines). The test portion of Fig. 1 might also be summarized as showing that: (*a*) when training treatment is constant, the ranking fits level of restriction in the test period (compare groups with the same connecting line and note that open circles > Xs > closed circles); (*b*) when test treatment is constant, the effects are as

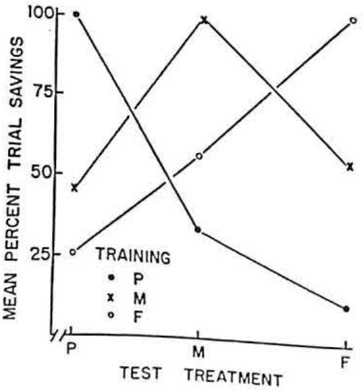


FIG. 2. Mean percentage of trial savings as a function of test treatment, training treatment the parameter.

consistent, but require more lengthy description; and (*c*) there is a great deal of transfer of training from one path width to another, but far from 100% transfer.

An analysis of ranked data (Kruskal & Wallis, 1952) based on *S* totals of Day 7 was used to test the significance of the effect of training treatment on test performance, for each test treatment separately. For P-, M-, and F-test treatments, respectively, H (2) = 29.1, 20.6, and 18.0, $p < .01$.

Intertreatment transfer.—Figure 2 plots percentage of transfer on the first test trial as a function of test treatment. Training treatment is the parameter. The test treatments are located along the abscissa in proportion to tolerance in handle turning. The points above Treatment P are from the data of Exp. I, the other points from the data of Exp. II. The number of trials saved was taken as the ordinal number of the control trial whose mean was closest in magnitude to the mean of the first test trial. Group P-P supplied the control data in Exp. I; Groups M-M and M-F served as a combined control comparison for transfer to Treatment M, and Groups F-F and F-M as a combined control comparison for transfer to Treatment F. For Groups P-P, M-M, and F-F, transfer is 100%; for the other groups transfer in Fig. 2 is expressed as the percentage of trials (out of 31) saved. For the extreme treatments, P and F, transfer decreases as the difference in training and test treatments increases. For the intermediate treatment, transfer decreases with a training-test difference in either direction.

The curves of Fig. 2 thus resemble generalization gradients. The considerable symmetry evident in Fig. 2 is consistent with the general position that bidirectional gradients of generalization underlie many transfer-of-training phenomena (Osgood, 1949). That is not to say that we are dealing with response generalization for a single response, since *S* has a range of responding for even a single segment of the path and since there is doubtless some overlap in the response distributions of the treatments. The findings do clearly indicate that, in this *S*-paced task, test performance varies with training treatment, regardless of the direction of the target size (or precision) difference from training to test.

No evaluation of the minor suggestions of differential transfer in the measures used in Fig. 2 will be attempted. The *n*'s are small, the comparison cuts across experiments, the direction is opposite to expectation in *E*-paced tasks, and amount, rather than direction of shift in treatment is at least the chief determinant of transfer of training. If transfer both *to* and *from* P is somewhat out of line, this can be explained by the stimulus discrepancy between P and the other treatments. Treatments M and F are guided only by visual stimuli; Treatment P requires an auditory click as well. One could make a strong case that P-trained *S*s have acquired not only habits that lead to positive transfer in M and F treatments, but a habit (waiting for the click) that introduces an element of negative transfer. Figure 2 is scaled according to response differences, while stimulus and response differences do not perfectly covary. A final reason for restraint is the number of questions that can be raised about the measures used to compare transfer benefits in tasks differing in both absolute score and in trend (Gagné, Foster, & Crowley, 1948; Gibbs, 1951).

REFERENCES

BILODEAU, E. A. A further study of the effects of target size and goal attainment upon the development of response accuracy. *HumRRO res. Bull.*, 1952, No. 52-7. (a)

BILODEAU, E. A. A preliminary study of the effects of reporting goals as a function of different degrees of response accuracy. *HumRRO res. Bull.*, 1952, No. 52-4. (b)

BILODEAU, E. A. Accuracy of response as a function of target width. *J. exp. Psychol.*, 1954, **47**, 201–207.

BILODEAU, E. A. Motor performance as affected by magnitude and direction of error contained in knowledge of results. *J. Psychol.*, 1955, **40**, 103–113.

BILODEAU, E. A. Studies of target size and the control of psychomotor behavior through systematic transformation of knowledge of results. In G. Finch & F. Cameron (Eds.), *Symposium on air force human engineering, personnel, and training research.* (NAS-NRC Publ. No. 455) Washington: National Academy of Sciences-National Research Council, 1956. Pp. 17–24.

BILODEAU, I. McD., & BILODEAU, E. A. Transfer of training and physical restriction of responses. *Percept. mot. Skills*, 1958, **8**, 71–78.

GAGNÉ, R. M., FOSTER, H., & CROWLEY, M. E. The measurement of transfer of training. *Psychol. Bull.*, 1948, **45**, 97–130.

GIBBS, C. B. Transfer of training and skill assumptions in tracking tasks. *Quart. J. exp. Psychol.*, 1951, **4**, 99–110.

KRUSKAL, W. H., & WALLIS, A. Use of ranks in one-criterion variance analysis. *J. Amer. Statist. Ass.*, 1952, **47**, 583–621.

MELTON, A. W. (Ed.) Apparatus tests. Washington: United States Government Printing Office, 1947. (*AAF aviat. Psychol. Program res. Rep.*, No. 4.)

OSGOOD, C. E. The similarity paradox in human learning: A resolution. *Psychol. Rev.*, 1949, **56**, 132–143.

(Received April 23, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 2, 141–146

RELATIONSHIP BETWEEN STATIC AND DYNAMIC VISUAL ACUITY[1]

SEYMOUR WEISSMAN[2] AND C. M. FREEBURNE

Bowling Green State University

Research in the area of dynamic visual acuity (DVA) has pointed out a controversy as to the nature of the relationship between DVA and static acuity. This study tried to answer the following questions: (a) Is there a relationship between static acuity and DVA at any speed? (b) If there are relationships at different speeds, are they different kinds of relationships? 30 women, college students, were given 6 speed (20, 60, 90, 120, 150, and 180°/sec) and 1 static measure of acuity. Thresholds for the first 4 speeds were found to show a significant linear relationship with the static acuity thresholds. The relationship disappeared at the 2 highest speed thresholds.

The term "dynamic visual acuity" (DVA) is used to designate the ability of an *O* to discriminate an object when there is relative movement between *O* and the object. Miller and Ludvigh studied this visual ability extensively (Ludvigh, 1949; Ludvigh & Miller, 1958; Miller, 1958). Their results can be summarized as follows:

1. Acuity for a moving target deteriorates markedly and progressively as the angular velocity of the target increases.
2. DVA performance can be improved through increased target illumination.
3. The two above findings apply substantially whether the target movement is horizontal or vertical or whether the target is moving with *S* stationary or vice versa.

In addition, there has been a controversy concerning the relationship between static visual acuity and DVA. Ludvigh and Miller (1958) found that there was a very low correlation between DVA and static acuity, and individuals with similar static acuity may differ significantly in their dynamic visual acuity. On the other hand, Burg and Hulbert (1961) noted a significant relationship between DVA and static acuity thresholds.

Besides the discrepancy in findings of the above studies, both investigators failed to test for possible nonlinear relationships between DVA and static acuity. By omitting a test for linearity, the previous investigators failed to test thoroughly the independence of these two types of visual-discrimination abilities.

The object of this study was to investigate the relationship between DVA and static acuity at a wide range of speeds.

METHOD

Subjects.—The *S*s were 30 undergraduate and graduate women. Their ages ranged from 18 to 40 but 22 of the *S*s were 20–25 yr. of age. Twenty-one of the *S*s wore their corrective lenses, since they used them for discriminating objects at a great distance.

Apparatus.—The apparatus consisted of a Kodak 550 Carousel slide projector mounted on a revolving turntable, which projected a target image on a 180° semicircular screen which was coated with white linoleum. The radius was 4 ft. Illumination for the testing room was provided by three 100-w. bulbs set in the ceiling of the room. The screen was illuminated by a 500-w. bulb in the projector. The *S* sat underneath the projector, so the pivot point of the turntable, the center curvature of the screen, and *S*'s head were all in vertical alignment. The angular velocities that were used included 20, 60, 90, 120, 150, and 180°/sec. The projector had a shutter device to limit exposure for DVA and static acuity projection to 1 sec. The shutter was a No. 5 Universal manufactured by Ilex Optical Company.

---

[1] This article is based on a thesis submitted by Seymour Weissman in partial fulfillment of the requirements for the degree of Master of Arts at Bowling Green State University.

[2] Now at United States Naval Medical Research Laboratory, Groton, Connecticut.

The targets for the DVA and static acuity measurements, were seven Landolt rings. The rings were randomly presented in each of four positions, i.e., with the opening at the left, right, bottom, and top, from *S*'s viewpoint. The following angles of the gaps (given in minutes) were used: 3.35, 2.854, 2.374, 1.894, 1.414, 0.934, and 0.454.

To produce the Landolt rings, seven drawings 20 times larger than the original dimensions were constructed. Following that, photographs were taken of the drawings to reduce their size to the original dimensions stated above.

Procedure.—The *S*s were divided into two groups of 15 each. The order of administration of the tests to the two groups was counterbalanced. Group 1 received the static acuity measure first, followed by the DVA measures. The *S*s in Group 2 were given the tests in the reverse order.

For the DVA and static acuity measurements, *S*s were tested binocularly and were allowed free head movement. The targets were shown to *S*s in random size order and were randomly positioned for each of 10 trials. The *S* was also tested for 10 trials for each angular velocity. Each testing session took approximately 2 hr., and three 5-min. breaks were taken during each session.

TABLE 1

INTERNAL CONSISTENCY RELIABILITY COEFFICIENTS

	r
Static acuity	.974**
20°/sec	.950**
60°/sec	.920**
90°/sec	.880**
120°/sec	.871**
150°/sec	.797**
180°/sec	.842**

** $p < .01$.

TABLE 2

CORRELATIONS BETWEEN STATIC AND DYNAMIC VISUAL ACUITY THRESHOLDS

	r
SVA & 20°/sec	.713**
SVA & 60°/sec	.675**
SVA & 90°/sec	.638**
SVA & 120°/sec	.665**
SVA & 150°/sec	.231
SVA & 180°/sec	.098

** $p < .01$.

Before each DVA and static test, *S* was shown a practice target to familiarize her with the type of object and the speed and path of the targets to follow. For the DVA tests, each target made only one sweep across the screen with the projector carriage shifting automatically to the next slide after the target had left the screen. There was a 10-sec. interval between slides for both DVA and static acuity tests. The *E* started the turntable, and once the target encountered the screen, the shutter was manually released, and the target remained on the screen for 1 sec. One second before each target appeared on the screen, the word "Ready" was given by *E* as a ready signal for *S*. The *S* was required to respond "top," "bottom," "right," or "left" to each target and to guess if undecided.

The static acuity test and target presentation was essentially the same as used for the DVA tests.

RESULTS

The proportion of correct responses for each gap size was tabulated. A correction for guessing using the formula $\frac{\text{Proportion Correct} - .25}{.75}$ was applied to each proportion (Dember, 1963). These corrected proportions were then used to compute the visual-acuity thresholds.

To test the reliability of the measuring instruments, split-half correlations were computed for each of the seven measures of acuity. These coefficients were obtained by correlating correct responses between odd and even trials. The internal consistency coefficients can be found in Table 1. All seven internal consistency reliability coefficients were significant at the .01 level.

The correlations between *S*s' static acuity thresholds and DVA thresholds at each of the six speeds are presented in Table 2. At all but the two fastest speeds (150 and 180°/sec) relationships between DVA and SVA were significant (.01 level). It should be noted that there is a decrease in the magnitude of the coefficients of correlation as the speed is increased. The one exception to this trend occurs at the 120°/sec measurement.

To determine better the nature of the relationships between DVA and static acuity thresholds, tests of linearity were applied to these data. For this purpose, the statistic $t = \frac{b}{s_B}$ was employed (Huntsberger, 1961). Table 3 presents these statistics. The thresholds at the four lowest speeds showed linear relationships with static acuity thresholds at the .01 level of significance, while the thresholds at the two fastest speeds (150°/sec and 180°/sec) failed to show significant linear relationships, either at the .01 or .05 levels.

On the basis of the preceding results, it seemed desirable to explore the nature of the relationship between thresholds at the two fastest speeds and the static acuity thresholds. The correlation ratio, η, was therefore computed to determine whether or not there was a nonlinear association between the 150°/sec and 180°/sec thresholds and the static acuity thresholds. The results of these computations are presented in Table 4. Both the 150°/sec and 180°/sec thresholds show nonlinear relationships with the static acuity thresholds. In Tables 3 and 4 the degree of linear or nonlinear association between the variables has been presented. To give a complete picture of the relationship between the DVA and static acuity thresholds, the form of the relationship must be described. These relationships are presented in Fig. 1. From these graphs, it can be seen that the first four DVA thresholds (20, 60, 90, 120°/sec) take the form of a linear relationship. The graphs of the 150 and 180°/sec thresholds do not seem to show a simple curvilinear relationship with the static acuity thresholds, even though the relationship is not linear.

The mean thresholds (in reciprocal minutes of arc) and their corresponding standard deviations at both static and angular velocities are presented in Table 5.

TABLE 3

t TESTS OF LINEARITY BETWEEN DVA AND STATIC ACUITY THRESHOLDS

	t
SVA & 20°/sec	5.37**
SVA & 60°/sec	4.85**
SVA & 90°/sec	4.38**
SVA & 120°/sec	5.22**
SVA & 150°/sec	1.26
SVA & 180°/sec	.528

** $p < .01$.

TABLE 4

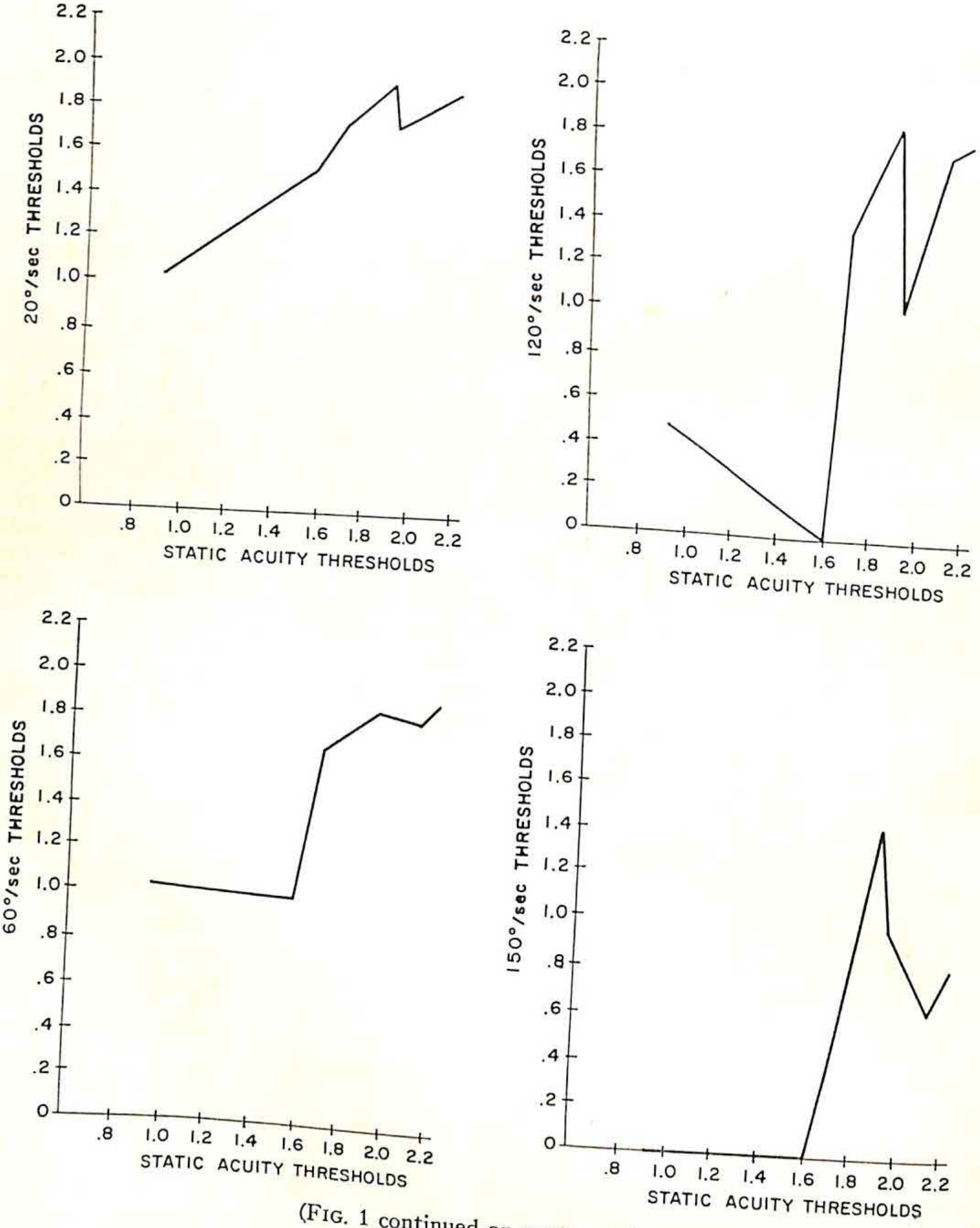
η AND F VALUES FOR 180°/SEC AND 150°/SEC AND STATIC ACUITY THRESHOLDS

	η	F
Static acuity & 150°/sec	.581	4.22*
Static acuity & 180°/sec	.582	4.43*

* $p < .05$.

DISCUSSION

This study supports, in part, the findings of the Burg and Hulbert (1961) study. They found a relationship between DVA and static acuity at all speeds (at the .01 level of significance), which were the same speeds as used in the present study. The present study found a significant relationship at the .01 level in all but the two fastest speeds. One possible reason for the discrepancy in results could be the difference in targets between the studies. Burg and Hulbert used orthorater targets for their static acuity measurements. Another possibility was the different populations used in the two studies. Burg and Hulbert (1961) used men and women motorists, while the present study used college women students.



(FIG. 1 continued on next page.)

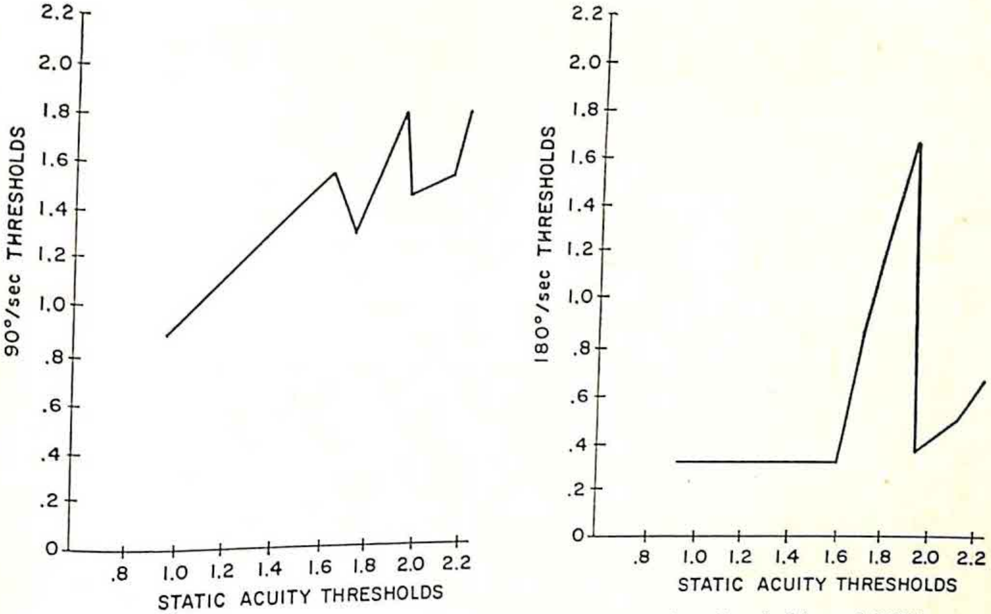


FIG. 1. Curves showing relationships between static acuity thresholds and DVA thresholds at six speeds (20, 60, 90, 120, 150, and 180°/sec).

Ludvigh and Miller found that there were no significant correlations, at either .05 or .01 levels, between DVA and static acuity at any speeds. The present study found a significant linear relationship at the .01 level for all but the two fastest speeds. Here, again, there are possible reasons for the discrepancies in the findings of the two studies. First, Ludvigh and Miller used Snellen ratings for static acuity, and DVA and static acuity were compared monocularly instead of binocularly. Second, Ludvigh and Miller used naval aviation cadets with 20/20 vision or better (uncorrected).

It should also be noted that the findings in Table 5 are comparable to those stated previously in the literature. In other words, acuity for a moving target deteriorated progressively as the target velocity increased. The variability of these thresholds increased in like manner as the speed increased.

The results of this study give partial support to both sides of the controversy. A significant relationship at the .01 level between the first four speeds and static acuity was found to exist, whereas at the two fastest speeds the relationship at this criterion level of significance was not demonstrated. It may be that the population sampled in this type of experiment is an important factor to consider. With a restricted range of static acuity thresholds, DVA thresholds at a wide range of speed would tend to deviate greatly from SVA scores. On the other hand, at a wide variety of SVA thresholds, the DVA thresholds obtained at the various speeds would tend more to be related to the static acuity thresholds.

TABLE 5

STATIC AND DYNAMIC VISUAL ACUITY THRESHOLDS CALCULATED IN RECIPROCAL MINUTES VISUAL ANGLE

	M	σ
SVA	2.04	.27
20°/sec	1.95	.29
60°/sec	1.84	.30
90°/sec	1.78	.32
120°/sec	1.63	.44
150°/sec	.90	.58
180°/sec	.94	.61

REFERENCES

Burg, A., & Hulbert, S. F. Dynamic visual acuity as related to age, sex, and static acuity. *J. appl. Psychol.*, 1961, **45**, 111–116.

Dember, W. N. *The psychology of perception.* New York: Holt, Rinehart, and Winston, 1963.

Huntsberger, D. V. *Elements of statistical inference.* Rockleigh, N. J.: Allyn & Bacon, 1961.

Ludvigh, E. *Arch. Ophthalmol.*, 1949, **42**, 14–22. (Abstract)

Ludvigh, E., & Miller, J. W. Study of visual acuity during the ocular pursuit of moving test objects: I. Introduction. *J. Opt. Soc. Amer.*, 1958, **48**, 799–802.

Miller, J. W. Study of visual acuity during the ocular pursuit of moving test objects: II. Effects of direction of movement, relative movement, and illumination. *J. Opt. Soc. Amer.*, 1958, **48**, 803–808.

(Received April 13, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 2, 147–150

EFFECT OF UCS INTENSITY UPON THE ACQUISITION OF CONDITIONED RESPONSES ACQUIRED UNDER A LENGTHENED INTERSTIMULUS INTERVAL[1]

KENNETH R. BURSTEIN

University of Massachusetts

Earlier investigations, using a single UCS intensity value, have shown that CRs acquired under a lengthened CS-UCS interval are independent of those conditioned at the original CS-UCS interval in terms of their latencies. A 2 × 2 factorial study in which pre- and postshift UCS intensities were orthogonal to each other confirmed previous results except in the case of a low preshift UCS intensity. An explanation for this discrepant finding is offered.

A previous study (Boneau, 1958) demonstrated that, after lengthening the interstimulus interval, the conditioned response appropriate to the initial interval gradually extinguished, while, at the same time, the level of responding to the lengthened interval increased in the same manner as it increased in groups conditioned at that interval with no previous conditioning experience. It was concluded that the response conditioned to a lengthened interstimulus interval is acquired independently of the response previously conditioned to the original interstimulus interval.

As Boneau points out, a fairly large percentage of *S*s were discarded for failure to give UCRs on five or more trials, presumably because of the weak air puff employed. The purpose of this study was to explore the role of UCS intensity in the acquisition of a CR acquired under a lengthened interstimulus interval.

METHOD

Apparatus.—The eyelid-conditioning apparatus previously employed at Duke University (Dufort & Kimble, 1958) was used in this experiment. The *S*s were seated in a sound-deadened, well-lighted cubicle before a tablelike shelf. Approximately 1 ft. above this shelf was another tablelike shelf upon which was a plywood reading stand large enough to accommodate the larger popular magazines.

The CS was a tone of 1,000-cps about 40 db. above threshold delivered through an overhead speaker. The UCS was an air puff of either 1.0 or 2.4 psi released by a solenoid valve and delivered from approximately $\frac{1}{2}$ in. to the temporal aspect of *S*'s right cornea through a nozzle with a $\frac{3}{64}$-in. aperture. The duration of the CS was either .55 sec. or 1.05 sec. and the duration of the UCS was .05 sec. with the CS and the UCS terminating simultaneously. Intertrial intervals, programed by a motor-operated disc, repeated the sequence 22, 14, 21, 20, 15, and 28 sec.

Subjects.—The *S*s were undergraduate Duke students participating in order to fulfill a course requirement. Of the 131 *S*s, 6 were discarded because of equipment failure, 1 was replaced due to a procedural error, 2 were rejected as voluntary responders according to criteria suggested by Spence and Ross (1959), and 2 were replaced for matching purposes. The remaining 120 *S*s were assigned haphazardly to one of eight groups, four experimental and four control.

Procedure.—There were 60 *S*s in the experimental conditions. For the first 50 trials these *S*s were conditioned with a .5-sec. interstimulus interval, half with the low UCS intensity and half with the high UCS intensity. In Trial 51, the interstimulus

---

[1] This experiment represents a portion of a dissertation submitted to the Psychology Department of Duke University in partial fulfillment of the requirements for the PhD degree. The author wishes to acknowledge his indebtedness to Gregory A. Kimble for advice and assistance throughout the course of the investigation.

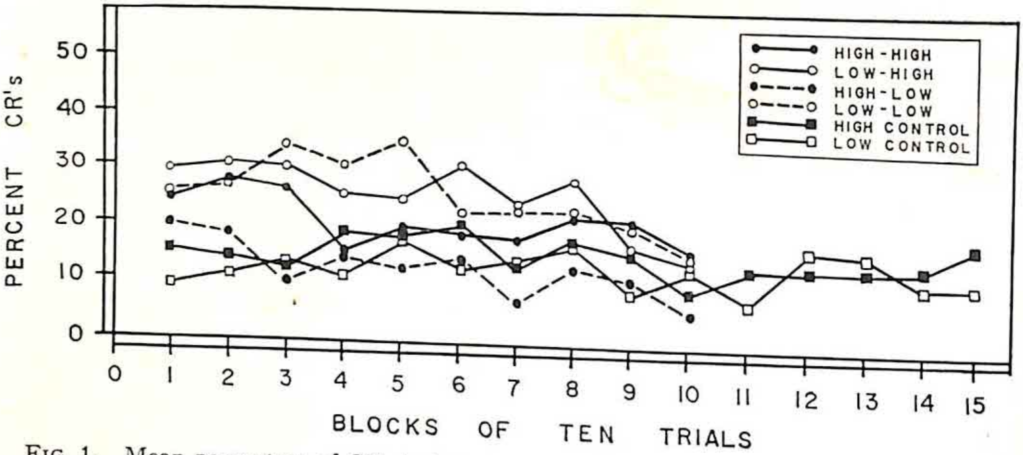


FIG. 1. Mean percentage of CRs in latency range .51–.65 sec. for the four experimental groups and for the 1.0-sec. control groups by blocks of 10 trials.

interval was lengthened to 1.0 sec. for 100 further trials and the UCS intensity was switched to the alternative value for half the *S*s in each group. Thus, in terms of UCS intensity employed before and after lengthening the interstimulus interval, there were four experimental groups: low-low, low-high, high-high, and high-low. Four control groups, each consisting of 15 *S*s, were designated the .5-high-control, the .5-low-control, the 1.0-high-control, and the 1.0-low-control groups. These groups received 150 trials at their appropriate UCS intensity and interstimulus interval. All *S*s were instructed to read magazines from an ample supply provided and to try to forget about blinking.

Definition of a CR.—Following the procedure adopted by Boneau (1958), a CR for the .5-sec. interstimulus interval was any deflection from the base line of 1 mm. or more having a latency of .31–.50 sec. Similarly, a CR for the 1.0-sec. interstimulus interval was any deflection from the base line having a latency of .66–1.0 sec. Those CRs falling between these two latency ranges, i.e., having a latency of .51–.65 sec., are the primary concern of this study.

RESULTS

All groups conditioned with the .5-sec. interstimulus interval attained a level of approximately 70% CRs at the end of the 50 trials. Except for minor random excursions this level was maintained in the (control) groups which received an additional 100 trials at the same interstimulus interval. All groups conditioned with a 1.0-sec. interstimulus interval reached approximately 45% CRs by the fiftieth trial at this interval and maintained this performance level until the experiment was terminated.

Figure 1 presents the mean percentage of CRs in the intermediate latency range by blocks of 10 trials for the four experimental groups and for the two 1.0-sec. control groups. An analysis of variance of the intermediate latency responses of the experimental groups reveals no significant effect of postshift UCS intensity, $F\ (1, 56) = 1.33$. The preshift intensity effect is significant at the .005 level, $F\ (1, 56) = 9.42$. As shown in Fig. 1, there is almost no overlap in the intermediate responses of groups conditioned with different UCS intensities for the first 50 trials. As Boneau points out, if the distribution of response latencies simply "slides" through the region intermediate between the two which define CRs, there should be a greater level of responding in the intermediate latency range for the experimental groups following the interstimulus interval shift than is exhibited by the control groups for the corresponding trials. Moreover, as trials continue, the initial increase in intermediate responding should be followed by a decrease as the latency first enters and then leaves the intermediate region. The trial effect for the experimental groups is significant at the .001 level, $F\ (9, 504) = 6.16$. A separate analysis of variance of the intermediate CRs of the two 1.0 control groups revealed no significant effect of trials, $F\ (9, 252) < 1$, or UCS intensity, $F\ (1, 28) < 1$. Another analysis of variance was performed on the intermediate CRs of the high-high experimental group and the 1.0-high-control group to determine whether the former group, although inferior in level of intermediate CRs to the low UCS intensity experimental groups, displayed a higher level of responding than the unshifted latter group, i.e., to assess the effect of interstimulus interval shift, per se. No significant difference attributable to UCS intensity, $F\ (1, 28) = 1.76$, or to trials, $F\ (9, 252) = 1.90$, was found. A parallel comparison of the low-low experimental group and the 1.0-low-control group revealed highly significant differences for UCS intensity, $F\ (1, 28) = 11.38$, and for trials, $F\ (9, 252) = 2.79$. Each of these F ratios is significant at the .005 level.

DISCUSSION

The results of this study indicate that, under certain conditions, CRs acquired under a lengthened interstimulus interval are not independent of those conditioned at the original interstimulus interval. While these results support such an independence of habits in respect to conditioning involving a strong UCS in the original interstimulus interval, they present quite a different picture with respect to conditioning involving a weak UCS in the original interstimulus interval. In the latter case it would seem that the response latency gradually shifts from its original value to one appropriate to the lengthened interval. This finding would support the results of Rodnick (1937).

While it is true that Boneau, using a lower UCS intensity than that employed in this study, found no significant difference in the intermediate CRs of his .5–1.0 shift group and his 1.0 control group, the former group did respond at a consistently higher level than the latter group, there being no overlap between the two groups. It is possible, therefore, that a Type II error may have been committed.

There is, of course, the possibility that these discrepant results stem from some procedural difference in the two studies. Boneau used a .5-psi air puff; this study employed a 1.0 psi air puff for the low UCS intensity. Thirteen of the 77 *S*s in Boneau's study were discarded because they failed to give five or more UCRs, presumably due to the weak air puff. Perhaps the weak air puff resulted in little or no conditioning in some *S*s which attenuated the effects reported. Some support for this interpretation is found in the fact that asymptotic performance for the .5 CS-UCS interval in Boneau's study is unusually low, being less than 40% CRs. There is also, of course, the possibility that the effect is related to UCS intensity in a nonmonotonic fashion.

Another procedural difference involves the activities of *S*s during conditioning. Boneau's *S*s were conditioned under fairly rigid conditions, i.e., they assumed a standard position with head placed on a chin rest, etc., while *S*s in this experiment were allowed to read magazines and to have much greater freedom of movement. Also, the CS in this study was a tone rather than the visual CS used by Boneau, since the latter did not seem as effective a stimulus as an auditory one in that it distracted reading.

There is an interpretation which would completely support Boneau's results. If the latency of the eyelid response distributes itself normally around the mean latency, then during conditioning the more extreme deviants from the mean, i.e., the longest latencies, would be "cut off" by the presentation of the UCS. Upon lengthening of the CS-UCS interval, these "latent" latencies would make their appearance as intermediate CRs. One possible explanation for the relative absence of these intermediate CRs in groups initially conditioned with a high UCS intensity may be related to the results of Gormezano and Moore (1962). These investigators found that the latency of the eyelid CR decreased as UCS intensity increased. If this finding reflects a shifting of the distribution of CR latencies without any appreciable change in dispersion, as postulated by Spence (1956), it would account for the differential performance of high and low UCS intensity groups, since most of the eyelid latencies would fall within the CS-UCS interval for the group conditioned with the high UCS intensity.

REFERENCES

Boneau, C. A. The interstimulus interval and the latency of the conditioned eyelid response. *J. exp. Psychol.*, 1958, **56**, 464–471.

Dufort, R. H., & Kimble, G. A. Ready signals and the effect of UCS presentations in eyelid conditioning. *J. exp. Psychol.*, 1958, **51**, 185–191.

Gormezano, I., & Moore, J. W. Effects of instructional set and UCS intensity on the latency, percentage, and form of the eyelid response. *J. exp. Psychol.*, 1962, **63**, 487–494.

Rodnick, E. Characteristics of delayed and trace conditioned responses. *J. exp. Psychol.*, 1937, **20**, 409–425.

Spence, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

Spence, K. W., & Ross, L. E. A methodological study of the form and latency of eyelid responses in conditioning. *J. exp. Psychol.*, 1959, **58**, 376–381.

(Received April 23, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 2, 151–153

EFFECT OF UCS STRENGTH ON GSR CONDITIONING: A WITHIN-SUBJECT DESIGN[1]

DELOS D. WICKENS AND GORDON B. HARDING
Ohio State University

Using a within-S design, 20 Ss had one CS (light or tone) associated with a strong UCS (2.5-ma. shock) for 10 trials, and the other CS (tone or light) associated with a weaker UCS of 1.5 ma. The interstimulus interval was 1 sec. and the response measured was the GSR. The data during training and extinction indicate a clear superiority of the strong over the weak UCS-associated CS. The results are contrasted with a comparable between-S design where differences failed to reach significance, and the use of a within-S design for the GSR is recommended whenever possible.

Although there is a rather large amount of evidence which indicates that there is a positive relationship between UCS intensity and the amount of eyelid conditioning (Ross & Hunter, 1959), Wickens, Allen, and Hill (1963) failed to obtain statistical support for any relationship between conditioned response magnitude and UCS strength in GSR conditioning. The procedure of that experiment employed a between-*S* design, wherein the CS was associated with a 2.5-ma. shock for one group and a 1.0-ma. shock for the other. With 40 *S*s in each group the difference, though favoring the strong shock group, reached a two-tailed significance level of only .16.

The general problem of the measurement of stimulus-intensity effects in the between- and within-*S* design has been considered by Grice and Hunter (1964), and they present considerable evidence to the effect that minor differences obtained in a between-*S* design are magnified in the within-*S* procedure. Support for this position with respect to CS strength is found in the work of Beck (1963). The present experiment was conducted in order to evaluate the effect of UCS intensity in GSR conditioning using the within-*S* design.

METHOD

Subjects.—A total of 32 *S*s was run, 12 being rejected, 3 because of *E* errors, 4 because of failure to condition, and 5 because of base levels above 100,000 ohms. Therefore 20 *S*s were used in the final analysis.

Apparatus.—The type of apparatus used has been described in detail elsewhere (Wickens et al., 1963). It consisted of a Fels dermohmmeter for recording the response, Tektronix equipment for timing, an Applegate (Model 350) shock supply, an oscillator, and a light source. The *S* and *E* were in adjacent rooms, *S*'s room being dimly illuminated and quiet.

Procedure.—Two conditioned stimuli were employed, a 1,000-cps tone of 55 db. re .0002 dyne/cm^2 and a light equated with the tone in perceived intensity. One of these was associated with a 1.0-ma. shock or UCS, the other with a 2.5-ma. shock. Ten *S*s experienced the strong UCS with the tone and the light with the weak and 10 experienced the reversed pairing. The interstimulus interval was 1,000 msec., the CS terminating simultaneously with the 100-msec. shock.

Following directions which informed *S* that one stimulus was going to be associated with a stronger shock than the other (though the modality was not identified) *S* was given three shocks alone of 1.5, 1.7, and 1.9 ma. Conditioning was then begun, and 10 trials were given to each CS-UCS pair. In the instance of the strong UCS the value was increased in successive trials in .2-ma. steps

---

[1] This research was supported by National Science Foundation Grant G19606 to the senior author.

TABLE 1

MEANS AND *SD*s IN LOG CONDUCTANCE UNITS FOR THE REINFORCED, NONREINFORCED, AND EXTINCTION TRIALS

Trials		CS + UCS		Training		Extinction	
		Strong	Weak	Strong	Weak	Strong	Weak
1	*M* σ	2.60 .29	2.53 .34	2.36 .43	2.11 .60	1.86 .73	1.57 .82
2	*M* σ	2.53 .35	2.41 .39	2.16 .74	1.99 .81	1.74 .71	1.55 .85
3	*M* σ	2.51 .39	2.43 .46	2.00 .57	1.56 .98	1.60 .96	1.41 .93
4	*M* σ	2.39 .47	2.17 .69				
5	*M* σ	2.27 .47	2.10 .63				

from 2.1 on the first trial to 2.5, at which value it remained.

Each CS alone was presented three times during the sequence of the 20 reinforced trials, and the experiment was terminated with 6 extinction trials, 3 to each CS. Order of testing to the CS associated with the weak or strong UCS was counterbalanced. The intertrial intervals ranged from 30 to 60 sec., with a 45-sec. interval preceding each CS test during acquisition and a 30, 45, and 60 interval being used for the 3 extinction trials to each stimulus.

The raw resistance data were recorded and converted to log conductance scores by the method and transformation described elsewhere (Wickens et al., 1963).

RESULTS

Table 1 presents the means and σ's for the data of the experiment. The sets of scores under the CS + UCS column represent the mean of the two adjacent reinforced trials of the indicated UCS value. Examination of this table shows that the magnitude of response to the weak stimulus is less than that to the strong in every pair, and also that there was a progressive decrement in response across trials under all conditions. A Trials × Group analysis of variance conducted on each set of data obtained the following results:

1. For the CS + UCS data, the effects of Intensity, $F\ (1, 16) = 13.89$, and of Trials, $F\ (4, 64) = 8.86$, were significant at the .01 level. There was also an Intensity × Trials interaction which was significant at the .05 level, $F\ (4, 64) = 2.60$. This interaction results from the somewhat steeper decline of response magnitude across trials for the weak compared with the strong stimulus.

2. For responses to the CS alone during training, the effect of Trials was significant at the .01 level, $F\ (2, 32) = 14.42$, and that of Intensity at the .05 level, $F\ (1, 16) = 6.12$. No interaction or other effects approached the .05 level.

3. For responses during extinction, the effect of Intensity was significant at the .05 level, $F\ (1, 16) = 6.36$, as was the effect of order, $F\ (1, 16) = 5.21$. The difference involving the order term arose as a result of the high level of performance and high resistance to extinction of the group tested to the conditioned stimuli in the order WSSWSW as opposed to the reverse order. The reason for the difference is not at all apparent and it seems likely that it should be attributed to chance.

Discussion

The results of the experiment very clearly demonstrate that the magnitude of the conditioned response is a function of UCS magnitude in GSR conditioning. It is apparent that the performance difference produced by the two conditioned stimuli cannot be attributed to a difference in drive since such a difference could not obtain in the within-*S* design.

Grice and Hunter (1964) have found that the within-*S* design produced a greater differential in mean response magnitude for CS intensity than did a between-*S* design, but it is also obvious that the error term would be smaller for the within-*S* design. It is pertinent to ask which of these variables is responsible for the higher level of confidence obtained in this experiment than in the Wickens, Allen, and Hill study. Because the differences in procedure in the two experiments involve more than a within- and between-*S* design the comparison between the experiments is only suggestive and cannot be made for all measures.

At one point in the two experiments the mean difference in response to the weak and strong CS is almost identical, and this is for the last two reinforced trials. The difference between the means for the within-*S* design is 99% the size of difference between the means in the between-*S* design, but the within-*S* error term is only 62% of the between-*S* error term. The fact that a $p < .01$ is obtained in one instance and $.10 > p > .05$ in the other can be attributed clearly to the smaller error term. There is, in addition, some hint that the within-*S* design produces a greater mean difference, for during extinction the between-*S* weak stimulus mean is 93% of the strong stimulus, while it is 87% of the strong stimulus in the within-*S* design. However, it would appear that the major factor responsible for the increased sensitivity of the within-*S* design is the reduction in the size of the error term.

It seems to the writers that efficiency in research with the GSR is usually hampered by a large error term arising from the wide range in base level found in the population and from the fact that the usual transformations of the response do not achieve complete independence of base level. By eliminating this source of variance, the within-*S* design in GSR research achieves a very considerable gain in efficiency. In the present experiment, for example, it is quite probable that stimulus generalization should have served to diminish the difference in response to the two CSs, and this could not occur in the between-*S* design. Despite this possibility, the within-*S* design produces more definitive results.

At the level of conditioning theory, this experiment may be used to support the view (Spence, 1956) that in classical conditioning the effect of UCS strength operates upon the intervening variable of habit (*H*), rather than upon another variable such as *K* which is used to handle incentive variations in instrumental learning. Another way of stating this is to say that in classical conditioning, because of its environmental simplicity, there is neither need nor justification to differentiate an intervening variable based on *N* from one based on incentive magnitude.

REFERENCES

Beck, S. B. Eyelid conditioning as a function of CS intensity, UCS intensity, and manifest anxiety score. *J. exp. Psychol.*, 1963, **66**, 429–438.

Grice, R. G., & Hunter, J. J. Stimulus intensity effects depend upon the type of experimental design. *Psychol. Rev.*, 1964, **71**, 247–256.

Ross, L. E., & Hunter, J. J. Habit strength parameters in eyelid conditioning as a function of UCS intensity. *Psychol. Rev.*, 1959, **9**, 103–107.

Spence, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

Wickens, D. D., Allen, C. K., & Hill, F. A. Effect of instructions and UCS strength on extinction of the conditioned GSR. *J. exp. Psychol.*, 1963, **66**, 235–240.

(Received April 24, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 2, 154–162

SHORT-TERM MEMORY UNDER WORK-LOAD STRESS[1]

ROBERT SEIBEL,[2] RICHARD E. CHRIST, AND WARREN H. TEICHNER

University of Massachusetts

Work-load stress was manipulated in terms of conditions which determine informational input rate and internal processing rate as factors which produce a breakdown in performance as a result of overloading short-term memory. The results suggest that breakdown will not result from high input rates, but rather from high internal processing rates.

Studies using information measures (e.g., Alluisi, Muller, & Fitts, 1957; Anderson & Fitts, 1958; Klemmer & Muller, 1953; Quastler, 1955) have shown that beyond some input rate, information transmission through *S* decreases with further increases in input rate. Similar results hold for data described by percent correct (e.g., Teichner, 1963; Teichner, Reilly, & Sadler, 1961). The usual interpretation is that *S*'s capacity to accept and hold data in short-term memory has been exceeded; the decrease represents a resultant breakdown of performance.

Both percentage and informational measures are relative measures in that both represent the degree to which *S*'s R output is related to the stimulus input. Thus, neither reflects *S*'s actual memory capacity. When the studies which have used these measures are examined with regard to the absolute number of correctly reported items, a performance breakdown is not consistently suggested. Rather, the number correct reaches a limit and tends to remain at that limit regardless of increases in input rate. Thus, either short-term memory has not yet really been studied at maximum capacity or it cannot be overloaded. The present study attempted to test the first possibility and, assuming it to be demonstrable, to ask whether a breakdown in absolute performance could be prevented by sacrificing a portion of the input information. That is, will *S* report more correct items when there is less to be reported than when there is too much?

EXPERIMENT I

Method

Apparatus.—Displays were projections of 160 negative slides containing randomly selected and positioned letters described by Teichner (1963). The number of letters on a slide, defined as the "load," was 4, 5, 6, or 7. Forty slides were made for each load. Instruction slides, reading "5th," "4th," "3rd," "2nd," or "1st" were also used. The projection screen was 12.5 ft. from *S*'s eyes; the projection area was 64 × 44 in.; the letters were 1.5 in. high, and the instruction characters were slightly more than 3 in. high.

Slide exposures were regulated by a solenoid-operated shutter mounted immediately in front of the lens of the projector. The changing of slides, exposure time, and interslide interval were controlled by a prepunched paper tape coupled with a 2-sec. pulser. The experimental room was in semidarkness. The *S*s sat at an IBM card punch and "keyed" their responses to each individual slide on separate IBM cards. The prepunched paper tape also controlled the IBM key punch, automatically changing cards, and automatically locking and unlocking the keyboard, thus controlling the time allowed for a report.

Subjects.—Six volunteer undergraduate students were used as *S*s. All six had demonstrated an ability of 35 net wpm on a standard typewriter and all had served in at least one similar type of experiment in the past. They were paid by the hour and, in addition, competed for prizes of $10.00 and $5.00 for the highest scores summed over all experimental conditions.

---

[1] This research was carried out under Contract N61339-1303, with the United States Naval Training Device Center, Port Washington, New York.

[2] Now at Pennsylvania State University.

Procedure.—The 40 slides of each load were divided into eight 5-slide sequences. Each slide was exposed for 2 sec., with 2 sec. between slides. Two seconds after a sequence of 5 slides an instruction slide was presented which told *S* on which slide he was to report. For the Full Report (FR) condition *S* reported on the "5th" (last one seen) slide first, then "4th," etc. until the last report was made on the first slide seen. Four seconds were allowed for keying the report of each slide, with 2 sec. between reports. A 6-sec. waiting period followed the last reporting period before a new stimulus sequence started. For the Partial Report (PR) condition there was only one instruction slide and one reporting period prior to the 6-sec. waiting period. The *S* was instructed to report on one of the middle three slides in the sequence, never the first or last. Thus, a 5-slide sequence was followed by one of four report conditions, namely: (*a*) FR; (*b*) PR calling for Slide 2; (*c*) PR, Slide 3; and (*d*) PR, Slide 4. The combination of a sequence of 5 slides plus one of the four report conditions will be referred to as a trial.

There were 8 trials under each load condition, 2 per report condition, and four different loads, yielding 32 trials per experimental session. There were five such sessions. Order was randomized within each load condition, and independently for each session. There were 5-min. resting periods between load treatments within a session and at least 24 but not more than 48 hr. between sessions.

The *Ss* were randomly assigned to two groups of three *Ss* each, and were run as a group whenever scheduling permitted. One group always received the four load conditions in ascending order within each session, and the other in descending order.

Prior to the main experiment all *Ss* were given two practice sessions with slides containing first one, and then two letters per slide, in five-slide sequences. All *Ss* attained near-perfect performance during these sessions.

Results

As the differences between groups were inconsistent and always <1.0 letters per slide, the groups were combined for analysis. Figure 1 presents the mean number of correctly reported letters per slide as a function of load for the FR and PR conditions. The straight, dashed lines represent perfect performance. The figure shows that for FR, performance decreased with the order of the slide reported. A reversal of this relationship is suggested at the fifth report. In no case did an FR curve increase and then decrease with increasing load. However, such a phenomenon is suggested for the PR at the second and fourth slides.

An analysis of variance of the FR data on which Fig. 1 is based did not indicate that load had a significant main effect, $F\ (3, 15) = 3.12$. Thus, the FR curves in Fig. 1 can be treated as random deviations from the horizontal. The main effect of position in the five-slide sequence, $F\ (4, 20) = 26.15$, and the interaction of load and position, $F\ (12, 60) = 7.36$ were significant ($p < .01$), the former supporting the decreasing trend across reports and the reversal on the last report shown in Fig. 1. The main effect of experimental sessions was also significant, $F\ (4, 20) = 13.56$ at $p < .01$. A plot of this effect showed a systematic increase in number correct of approximately .5 letters from the first to the last session. Nothing else reached significance at the .05 level except Load × Position × Session, $F\ (48, 240) = 1.69$.

An analysis of variance of the PR data shown in Fig. 1 yielded a significant effect of load, $F\ (3, 19) = 14.05$ and Load × Position, $F\ (6, 30) = 3.81$ at $p < .01$. Position, $F\ (2, 10) = 7.29$ and sessions, $F\ (4, 20) = 3.89$, were at $p < .05$. No other effect was significant at $p < .05$. Thus, this analysis supports the PR load effect shown in Fig. 1 and suggests that whether or not a bow-shaped curve will be

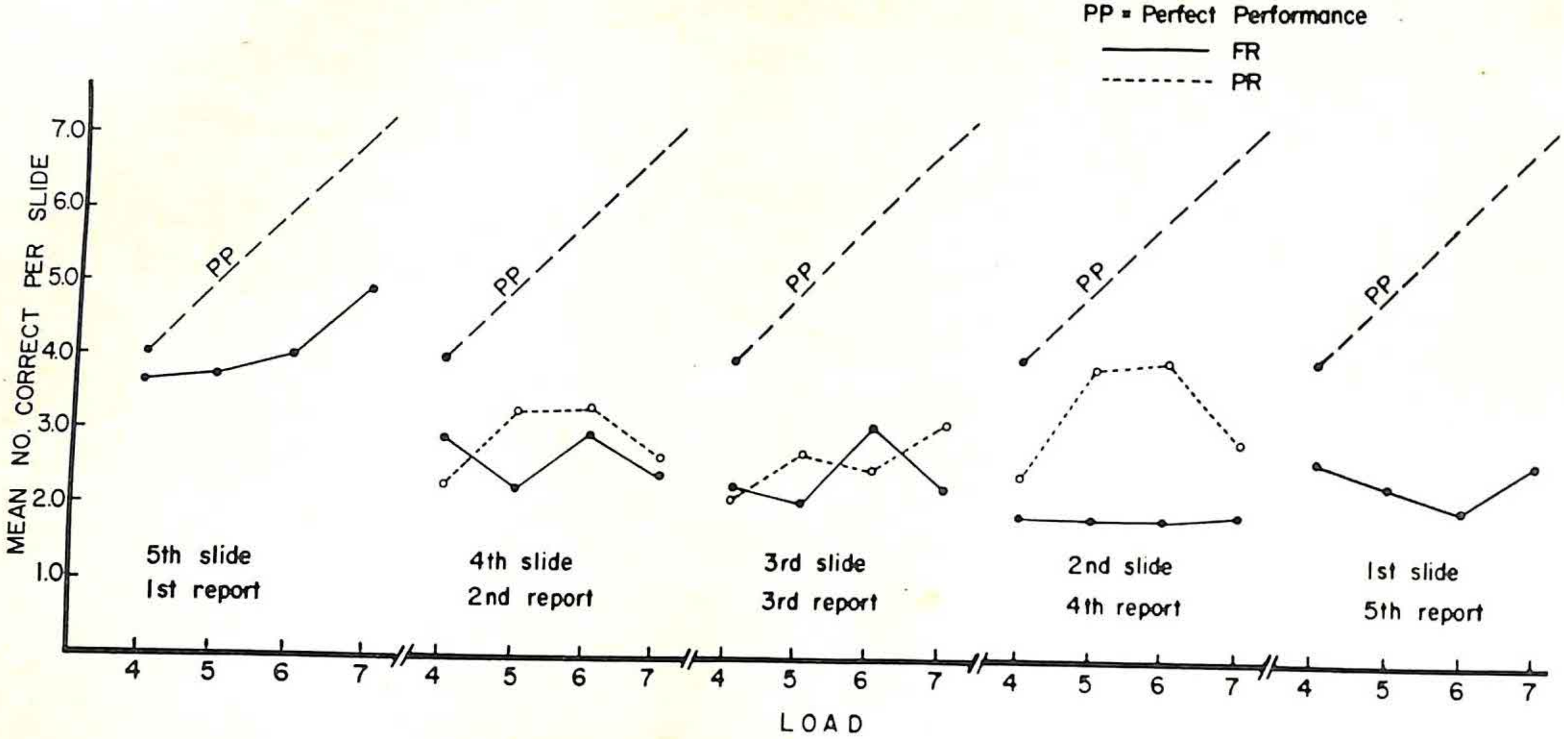


FIG. 1. Mean number correct as a function of load and position for the Full and Partial Report conditions in Exp. I.

obtained depends on the position in the series reported. The sessions effect, when plotted, was similar to that for the FR condition.

Figure 1 also suggests more correctly reported letters for the PR than FR conditions at the fourth and second slides. An analysis of variance of this comparison supported this interpretation by yielding a significant ($p < .01$) main effect for the reporting condition, $F\ (3, 15) = 11.09$.

EXPERIMENT II

Experiment II was designed to increase input load stress further by increasing the load per slide to a maximum of nine and by varying exposure time. Only FR conditions were used. Anticipating a performance breakdown, blank slides were substituted for some of the lettered slides to test the hypothesis that a reduction in input below those levels which produce breakdown will lead to a net increment in the number of correctly reported letters.

Method

Apparatus.—Additional slides were made, as required, to yield 90 slides for each of four load conditions, namely, six, seven, eight, and nine letters per slide. Blank slides were made by photographing a bright, diffuse source; the projected image was uniformly dark gray.

Subjects.—The six experienced *S*s of Exp. I served as before. Each *S* was run individually.

Procedure.—The general procedure was as in Exp. I. There were 60 trials per session, having four blocks of 15 for each of the four load conditions, with 10-min. rests between blocks. The load conditions were always given in ascending order. The 75 slides for each block were a random sample from the 90 available, and each *S* had a different set of random samples for each session.

Either zero, one, or two blank slides were inserted in a five-slide sequence, and these three number-of-blanks conditions were combined with two exposure times (1 or 2 sec.) to yield six treatment combinations. Each of the six *S*s was to take all six of these combinations according to a balanced Latin-square design in order to control for practice and first-order sequential effects. Blank slides were inserted only in the second, third, or fourth positions of the five-slide sequence, and position was randomized for each block of 15 trials with the restriction that each position appear equally often.

A 4-sec. reporting period was provided for each of the five slides in a trial, with 2 sec. between periods. For blank slides, *S* was instructed to wait during the appropriate reporting interval.

Results

The data of two *S*s who did not complete all six sessions were discarded. Thus, exposure time and number-of-blanks effects are partially confounded with sequential-order effects, but, assuming a linear practice effect for highly practiced *S*s, remain balanced with respect to practice effects when the data are collapsed across sessions.

An analysis of variance performed on the total number of letters correctly recalled per five-slide sequence indicated that exposure time was not significant, $F\ (1, 3) = 4.91$. The effects of load, $F\ (3, 9) = 32.05$, number of blanks, $F\ (2, 6) = 22.73$, and the Load × Blanks interaction, $F\ (6, 18) = 8.95$, were significant, each at $p < .01$. The other interactions were not significant at the .05 level.

Figure 2 presents the mean number correct per slide as a function of load with blanks and exposure time as parameters. The number correct increased as the load increased, especially for the zero-blank condition. For each level of load, an increase in the total number of slides which actually contained letters led to an increase in the number correct.

EXPERIMENT III

This experiment further increased input load stress by increasing the number of slides in the slide sequence

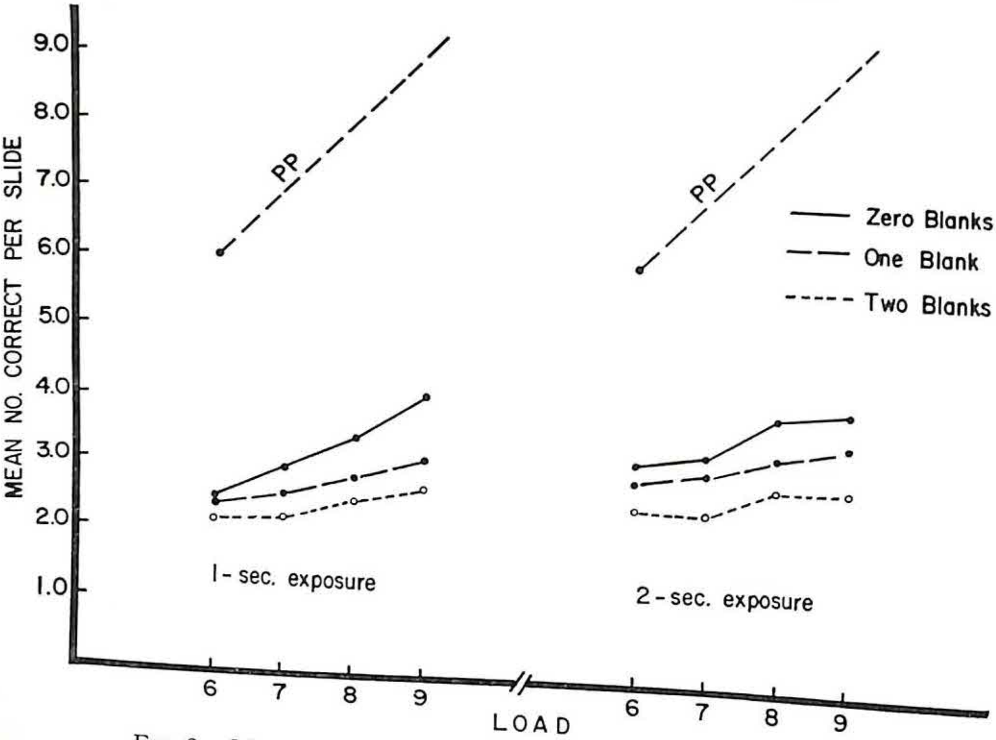


FIG. 2. Mean number correct as a function of load and exposure with number of blanks as a parameter for Exp. II.

(trial). In addition, Exp. III employed naive *S*s.

Method

Subjects.—Thirty students who had demonstrated a minimum typing speed of 35 net wpm were selected. They were paid by the hour and competed for cash awards for highest scores. Each *S* ran individually.

Procedures.—Sequences of five, seven, and nine slides were used. Ten *S*s, randomly assigned to each sequence, reported on all slides seen during each trial, keying nothing for report periods corresponding to blank slides. The load and blank-slide conditions were as defined in Exp. II. Each *S* responded for 36 trials in a single session. Three blocks of 12 trials each, corresponding to the three blank-slide conditions, were used, with order of blocks partially balanced over *S*s. Each block of 12 trials consisted of four sets of 3 trials each for each of the four load conditions, with load conditions administered in ascending order. Each one of the 3 trials corresponded to one of the three possible positions of blank slides, with order of trials independently randomized for each combination of variables, but the same for the 10 *S*s within each of the three length-of-sequence groups.

Slides were exposed for 1 sec. with 1 sec. between slides in a trial. Report periods were 4 sec. for each slide with 1 sec. between periods. There were rest periods of about 2 min. between successive load treatments and about 10 min. between successive blank conditions.

Prior to the main experiment all *S*s received three practice sessions with single slides, increasing in load as practice progressed. On the fourth session, just prior to the main experiment, each *S* was given experience with the length of sequence (and reporting procedure) to which he had been assigned.

Results

All analyses are in terms of the total number correct summed across the three-trial sets included in each combination of length, *S*, blank, and load. An analysis of variance indicated the following effects significant at the .01 level: length, *F* (2, 27)

$= 6.15$; load, $F(3, 81) = 40.08$; blanks, $F(2, 54) = 15.37$; Length × Load, $F(6, 81) = 3.26$; and Load × Blanks, $F(6, 162) = 5.79$. The Length × Blanks and the Length × Load × Blanks interactions were not significant at the .05 level. Figure 3 depicts the results summed over the 10 *S*s in each group. As load increased and as length increased so did the total number of correctly reported letters. In general, as number of blank slides increased the total number of correct letters decreased, but the effect was not as consistent as the other two main effects. The Length × Load interaction was primarily due to the more marked effect of the load on the performance of the Length 9 group. The significant Load × Blanks interaction is not very orderly.

DISCUSSION AND CONCLUSIONS

For both experienced and naive *S*s it was found that increasing the length of a series of slides, increasing the number of lettered slides per unit time, and increasing the number of letters per slide all led to an increase, or at least no marked decrease, in the total number of letters correctly reported when *S* was asked to make a full report of what he had been shown. For these conditions the results do not suggest that increases in input stress lead to breakdown.

Performance with PR did break down at two of the three slide positions. Although both FR and PR groups were comparable in regard to stimulus input conditions, they did differ in regard to the time available for extracting the information to be reported. The FR group was allowed 4 sec. of reporting time per slide in the five-slide sequence, whereas the PR group was permitted 4 sec. for the one slide it had to report. However, since the PR group was not told which slide to report until all slides had been presented, the 4-sec. period had to be used both to scan all slides and to report the items on the directed slide. As a result the internal processing rate of the PR group was greater than that of the

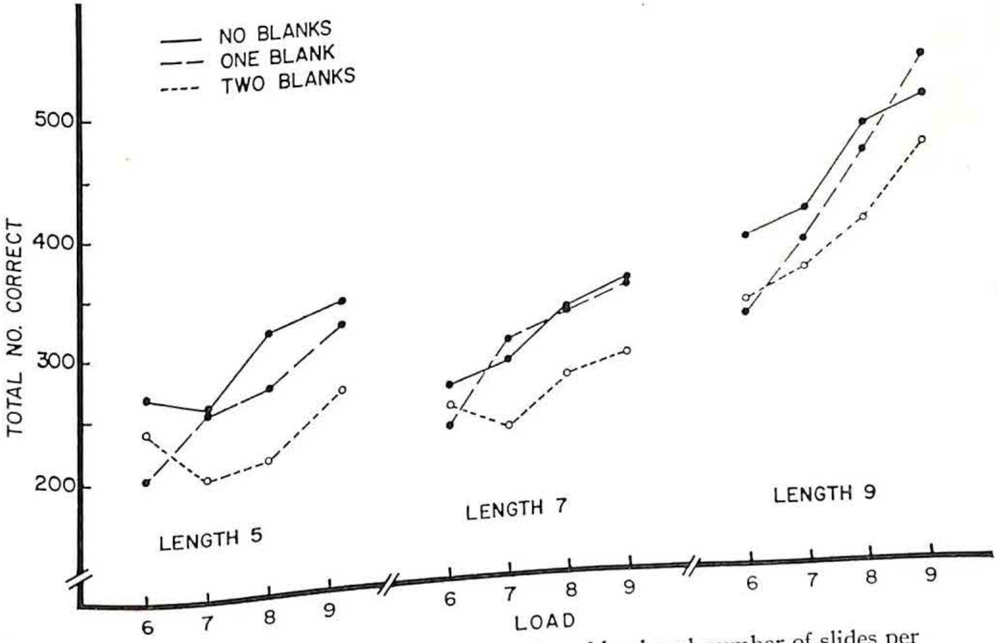


FIG. 3. Total number correct as a function of load and number of slides per report (length) with blanks as a parameter for Exp. III.

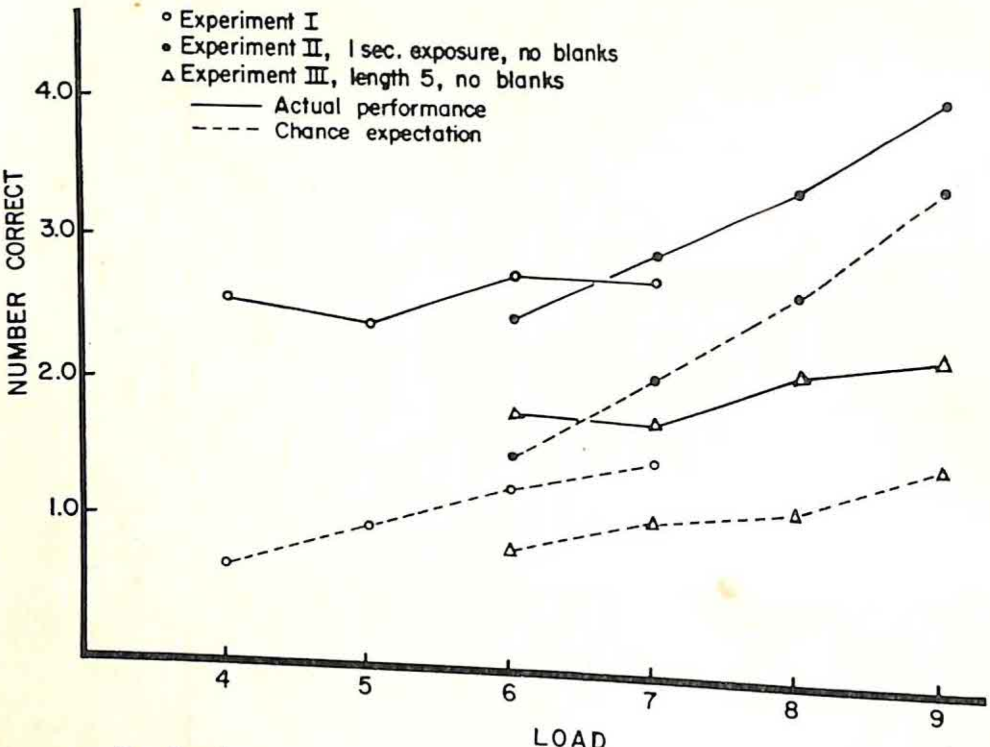


FIG. 4. Comparison of actual performance and predicted performance based on chance alone.

FR group. Thus, the results indicate that breakdowns are more likely to result from a high internal processing rate than from a high input rate. The failure of the PR group to exhibit a breakdown at the second slide position weakens, but does not invalidate this expectation. That result may have been experimental error or, as suggested by the significant Load × Position interaction, it may reflect a genuine position effect whose optimal condition may reduce the required internal processing rate. Intuitively, it does not seem unreasonable that *S* can locate the center position of a series more rapidly than off-center positions.

The expected number correct for any given load and for any given number of responses may be calculated via the hypergeometric distribution (Feller, 1957). Figure 4 depicts the comparison of actual number correct with the number expected correct by chance for certain comparable conditions across the three experiments. It may be seen that the four experienced *S*s of Exp. I and II tended to emit many more responses per slide than the naive *S*s of Exp. III. These differences in number of responses account for the differences in the expected numbers correct. In general, performance was above chance for all conditions, and exhibited a slight approach towards chance as load was increased. Thus, the data cannot be accounted for simply in terms of random guesses. The *S*s actually stored *something* in short-term memory. While the difference between actual number correct and chance number correct decreases slightly with increases in input work-load stress (load), there is no marked decrement which could be called breakdown.

Actual performance relative to chance suggests the hypothesis that *S*s filtered the input they were exposed to and accepted, on the average, only as much as they could handle, thus rendering the magnitude of work load ineffective

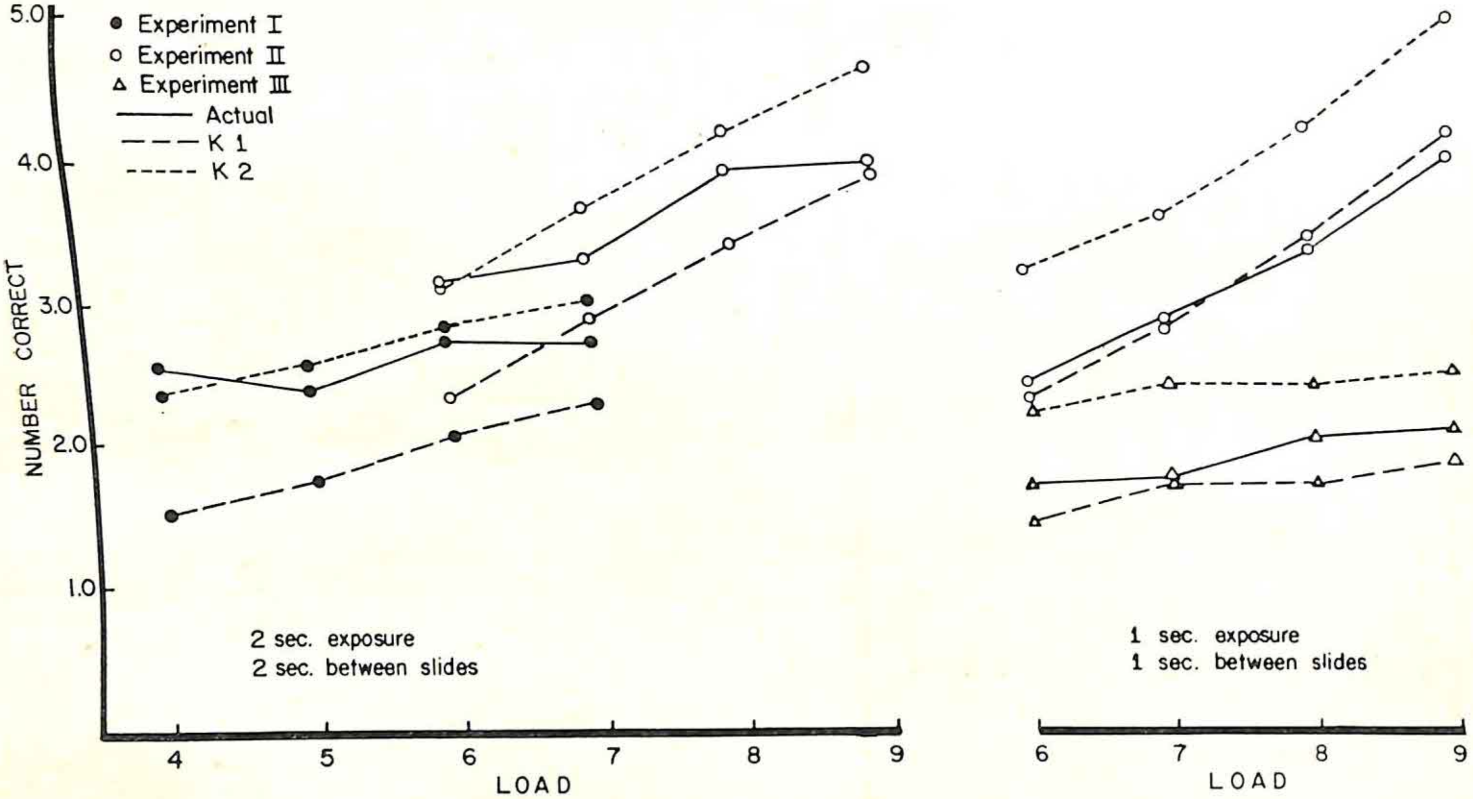


FIG. 5. Comparison of actual performance and predicted performance for K correctly stored letters, plus chance.

as a stressor. Assuming that *S*s actually stored one letter (or two) which they were able to report correctly, and then guess the rest of the letters that they actually reported, it is possible to generate, via the hypergeometric distribution, expected performance curves. Figure 5 depicts the expected and actual performance curves for comparable conditions from each of the three experiments. The curves labeled K1 assume one letter correctly stored, and those labeled K2 assume two letters correctly stored.

Figure 5 shows that a filter plus chance model accounts for the performance of the naive *S*s of Exp. III reasonably well. They stored (accepted), on the average, a little over one letter per slide, and guessed the rest of their responses. The amount stored was relatively constant despite increasing numbers of letters per slide.

The performance of the experienced *S*s of Exp. I and II differed. These *S*s showed systematic departures from the predictions of the model. It is as if the amount accepted by these *S*s systematically decreased with increasing load, and this amount was less overall for the shorter exposure time. Thus, in terms of a filter-type model, the results overall suggest a progressive narrowing of the filter with increases in load. On this basis, a performance breakdown would not be expected in terms of absolute measures of performance, but would appear in relative measures as an arithmetical result of relating a constant or decreasing R output to an increasing stimulus input.

In summary, the results of this study suggest that high input rates are ineffective as work-load stressors because *S* limits the amount of information he will accept at any one time. On the other hand, the results suggest that the rate at which *S* must extract information from memory is a variable which at high levels results in a breakdown in performance.

REFERENCES

ALLUISI, E. A., MULLER, P. F., & FITTS, P. M. An information analysis of verbal and motor responses in a forced-paced serial task. *J. exp. Psychol.*, 1957, **53**, 153–158.

ANDERSON, N. S., & FITTS, P. M. Amount of information gained during brief exposures of numerals and colors. *J. exp. Psychol.*, 1958, **56**, 362–369.

FELLER, W. *An introduction to probability theory and its applications.* (2nd ed.) New York: Wiley, 1957.

KLEMMER, E. T., & MULLER, P. F. The rate of handling information. Key pressing responses to light patterns. *USAF HFORL memo Rep.*, 1953, No. 34.

QUASTLER, H. *Information theory in psychology: Problems and methods.* Glencoe, Ill.: Free Press, 1955.

TEICHNER, W. H. Perception and short-term memory in the identification of multi-symbol displays. *J. exp. Psychol.*, 1963, **65**, 407–413.

TEICHNER, W. H., REILLY, R., & SADLER, E. Effects of density on identification and discrimination in visual symbol perception. *J. exp. Psychol.*, 1961, **61**, 494–500.

(Received April 25, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 2, 163–168

TACHISTOSCOPIC RECOGNITION THRESHOLDS, PAIRED-ASSOCIATE LEARNING, AND FREE RECALL AS A FUNCTION OF ABSTRACTNESS-CONCRETENESS AND WORD FREQUENCY[1]

WILMA A. WINNICK AND KENNETH KRESSEL

Queens College of the City University of New York

Comparisons were made of tachistoscopic recognition, paired-associate learning, and free recall by manipulating the variables of word frequency and the degree of abstractness of English words in each of these experimental situations. Concrete words were found to be more readily associated with nonsense syllables and to be better recalled than abstract words, but no such influence obtained for recognition thresholds. High-frequency words lowered thresholds and facilitated paired-associate learning, but had an unclear effect on free recall. In order to clarify some of the obtained relationships, measures of the meaningfulness of concrete and abstract words of high and low frequency were determined. Concrete words were found to have significantly higher meaningfulness scores than abstract; frequency showed a positive relationship to meaningfulness for the concrete words only.

Although the variables studied in tachistoscopic-recognition experiments have generally paralleled those that have proved effective in verbal learning, there have been few instances of parametric comparison of word recognition and verbal learning. One such comparison is provided in the recent study by Gibson, Bishop, Schiff, and Smith (1964) which measured tachistoscopic-recognition thresholds and two aspects of recall as a function of meaningfulness and pronunciability of nonsense syllables. The present study is a second experimental comparison, studying frequency of usage and degree of abstractness-concreteness of English words as the variables under scrutiny.

Four experiments were undertaken to compare recognition thresholds, paired-associate learning, and free recall as they are influenced by these two variables. The first variable, word frequency, has been studied in a large number of learning and recognition experiments, while abstractness-concreteness has been studied only in learning situations (Gorman, 1961; Stoke, 1929). It seemed possible that similarities or differences in the effect of these two variables on the experimental situations studied might aid in an analysis of the similarities or differences in the behaviors involved.

The first experiment measured tachistoscopic thresholds for mixed arrangements of abstract and concrete words of high and low frequency. In the second experiment, the same words were paired with nonsense syllables for paired-associate learning. The third and fourth experiments studied free recall for these words by two slightly different methods. The fourth experiment also measured the number of associations made to the four types of words in a 30-sec. time period.

---

[1] This study was supported by an undergraduate research grant (No. 2T2-7311-03) of the National Institute of Health and by a grant from the Graduate Division of the City University of New York. The first three experiments were summarized in a paper delivered to the 1964 meetings of the Eastern Psychological Association.

METHOD

Subjects.—The first two experiments used as *S*s undergraduate students drawn from a population of introductory psychology students required to participate in experiments. Experiments III and IV used as *S*s classes in experimental psychology. All *S*s were naive as to the purposes of the experiments.

Materials.—The experimental materials, identical in all experiments, were chosen from Gorman's (1961) list of abstract and concrete nouns and consisted of 32 words; of these, 16 were abstract, half of high and half of low frequency; 16 were concrete, half of high and half of low frequency. The abstract-concrete dichotomy was based on the empirical 2-point scale of Gorman (1961, p. 26). These words are shown in Table 1. For Exp. I and II, three lists were compiled of 16 words, four of each type; Lists 1 and 2 contained mutually exclusive sets of words; List 3 was composed of equal numbers of words in each category drawn from those of Lists 1 and 2. For Exp. II, the words, arranged by lists as described, were paired with nonsense syllables of 47% associative value (Glaze, 1928) to form nonsense-syllable word pairs for paired-associate learning. For Exp. III and IV, all 32 words were presented to *S*s. In Exp. III the words were printed on slides and projected to the group; in Exp. IV, a second group experiment, they were mimeographed on individual sheets of paper $4\frac{1}{4} \times 5\frac{1}{2}$ in. and arranged in booklet form.

TABLE 1

32 WORDS USED IN THE FOUR EXPERIMENTS

List 1	List 2	List 3
Abstract High Frequency		
OPINION (AA)	METHOD (AA)	LANGUAGE (A)
LANGUAGE (A)	PRINCIPLE (A)	PRINCIPLE (A)
GRATITUDE (22)	BIRTHDAY (37)	EXISTENCE (45)
EXISTENCE (45)	NONSENSE (19)	BIRTHDAY (37)
Concrete High Frequency		
TEACHER (AA)	STREET (AA)	FACTORY (A)
FACTORY (A)	HOSPITAL (A)	HOSPITAL (A)
CANDIDATE (45)	BEDROOM (35)	CANDIDATE (45)
BUTTERFLY (22)	ADMIRAL (20)	BEDROOM (35)
Abstract Low Frequency		
APPROVAL (14)	DIMENSION (10)	APPROVAL (14)
HEADACHE (10)	IDLENESS (9)	IDLENESS (9)
OCCURRENCE (9)	ANECDOTE (7)	OCCURRENCE (9)
SATIRE (7)	EPISODE (7)	EPISODE (7)
Concrete Low Frequency		
VOLCANO (14)	WATERFALL (10)	VOLCANO (14)
ORCHESTRA (10)	CALCIUM (9)	CALCIUM (9)
UNDERBRUSH (8)	FILAMENT (7)	UNDERBRUSH (8)
CANYON (7)	GELATIN (7)	GELATIN (7)

Note.—The number in parenthesis is the Thorndike-Lorge frequency count.

Apparatus.—For Exp. I, the words were typewritten in capital letters on cards for exposure in a Gerbrands tachistoscope. Experiment II presented nonsense-syllable word pairs in a long-arm Lafayette memory drum. In Exp. III, a projection tachistoscope was used to project slides made from the words. For Exp. IV, the words were individually presented on separate pages of a booklet and instructions and timing were tape-recorded.

Experimental design.—All experiments used a "mixed" arrangement of variables with three independent replications providing the sole "among" effect. In Exp. I and II, the three replications differed in lists learned and in Exp. III, in order of presentation of the 32 words. In Exp. IV, order was varied across *S*s within each replication, and replications were identical in all controllable respects. In all experiments, the dimensions of the materials—degree of abstractness-concreteness and word frequency—provided the "within" effects, all *S*s learning or being exposed to all four types of words.

Procedure.—In Exp. I, tachistoscopic thresholds were measured for each *S* for the words of one list only, following practice with sample words. Exposure intervals were begun at .01 sec. and increased by .01-sec. intervals until the word was identified correctly on any two, not necessarily consecutive, exposures. Each word was exposed five times at each exposure interval, providing as the threshold the number of such exposures required to reach two correct recognitions. Order of presentation of words in each list was randomized from *S* to *S*. Instructions were those used by Spielberger and Denny (1963).

In Exp. II, lists were learned by the standard anticipation method, with stimulus item exposed for 2 sec., followed by 2-sec. exposure of stimulus plus response item, and an 8-sec. intertrial interval. For each list five different arrangements were employed to prevent serial learning. Standard anticipation instructions were used, and learning was to a criterion of one errorless trial.

Experiment III was carried out as a group experiment, with the 32 words printed on slides and presented by a projection tachistoscope in three different orders for the three different groups. Exposure time was .10 sec. with approximately 1 sec. between exposures. Immediately following presentation of the words, four successive free-recall tests were given. To prevent rehearsal, response sheets were used which consisted of a plastic cover topping two data sheets separated by a carbon; using a stylus, *S*s wrote on these sheets without being able to see what was written.

Experiment IV was undertaken for two reasons: first, to determine whether the basis for the greater ease of associating concrete words with nonsense syllables was that such words of themselves have greater numbers of associations capable of mediating the required associations; secondly, to provide additional data on recall with order varying across *S*s. Accordingly, this was run as a group experiment; words arranged in different order for each *S* were printed one to a page in booklet form. The booklets were presented face down to *S*s, and they were instructed, by taped instructions, to turn over the booklet at the sound of the first buzzer and to write down as many associations as possible to the word on the first page, and at the sound of the next buzzer to turn the page and write the associations to the next word, etc. The interval between buzzer signals was 30 sec. Following this, booklets were collected, and the recall sheets described under Exp. III were passed out for a single recall test of 3-min. duration.

RESULTS

The data collected from the four experiments are summarized in Table 2; shown here are the means for the individual replications in each experiment and the overall means obtained for each of the four types of words used. Analysis of variance enabled, within each experiment, comparisons of main effects and interactions with the appropriate within-groups error terms.

TABLE 2

SUMMARY OF RESULTS OBTAINED FROM THE FOUR EXPERIMENTS

		Abstract Words		Concrete Words	
		High	Low	High	Low
Exp. I: Recognition thresholds (trials to criterion)	I	8.6	8.9	8.4	10.1
	II	10.9	11.7	10.5	11.2
	III	9.4	8.4	7.7	10.0
	M	9.6	9.7	8.9	10.4
Exp. II: Paired-associate learning (trials to criterion)	I	12.0	12.8	9.7	10.3
	II	11.7	14.8	8.3	12.0
	III	13.0	11.4	8.8	10.0
	M	12.0	13.0	8.9	10.8
Exp. III: Immediate recall (mean recall score)	I	2.9	2.8	4.2	3.6
	II	1.8	2.8	2.9	3.7
	III	2.4	2.8	3.7	4.0
	M	2.4	2.8	3.6	3.8
Exp. IV: Mean number of associations	I	7.6	7.8	9.6	9.2
	II	9.6	9.9	12.7	11.9
	III	10.6	10.6	12.5	11.6
	M	9.3	9.4	11.6	10.9
Exp. IV	I	2.6	2.7	4.7	4.1
	II	2.4	2.8	2.9	3.0
	III	2.6	3.1	4.4	3.0
	M	2.5	2.9	4.0	3.4

Threshold values.—The overall threshold for abstract words (measured in trials) was found to be identical with the measure for concrete words (9.65 for each). The values for low frequency were found to be greater than for high (10.05 and 9.25, respectively). The analysis of variance showed this to be the sole significant effect in this experiment, $F\ (1, 27) = 4.46, p < .05$. The main contribution to this effect derived from the concrete words, in contrast with the abstract where the difference between low and high frequency was almost negligible. The Type × Frequency interaction was short of significance, however.

Trials to learn paired associates.—The paired-associate learning data showed a clear influence of both variables studied, with concrete words and high-frequency words having the advantage. The analysis of variance showed the abstract-concrete dimension to be significant at the .01 level, $F\ (1, 27) = 30.05$, and word frequency at the .05 level, $F\ (1, 27) = 6.64$.

Free recall following tachistoscopic projection of words.—The four successive recall scores for each type of word were averaged for individual *S*s to form a single recall measure. The mean recall scores for the three replications (varying order), as shown in Table 2, show a clear effect of the abstract-concrete dimension, shown to be significant at the .01 level, $F\ (1, 27) = 21.25$, by analysis of variance. Although the main effect of frequency was not significant at an acceptable level, the interaction of frequency and replication was significant, $F\ (2, 36) = 11.39$, $p < .01$. Since order was varied in the three replications, this significant interaction indicates the possible effect of the ordering of words upon high- and low-frequency recall scores.

Meaningfulness of words.—Measures of the number of written associations to each of the stimulus words were made in the first part of Exp. IV; this measure corresponds operationally to Noble's (1952) meaningfulness (m) concept. The numbers of associative responses were averaged for each type of word for each *S*; the means for the three replications and the overall means are shown in Table 2. Analysis of variance showed the abstract-concrete dimension to have a highly significant effect, $F\ (1, 27) = 87.22$, $p < .01$, on the number of associations. The frequency effect was nonsignificant, but the interaction of the degree of abstractness and frequency was significant, $F\ (1, 27) = 27.40$, $p < .01$, with the frequency effect for concrete words contributing to this variance. Pursuing further this relationship, the product-moment correlation between the number of associations and the log frequency (L count) was found to be +.261, a value not significantly different from zero, $t\ (31) = 1.96$. When correlations were carried out separately for abstract and concrete words, the r for concrete words of .737 was found to be indicative of a significant positive relationship, $t\ (15) = 4.08$, $p < .01$; the correlation for abstract words was close to zero (obtained $r = .029$).

Free recall following association measurement.—In order to study the effect of frequency on recall with order of words varying across *S*s, a free-recall measure was taken following the measure of associations. Since in this experiment the measure of associations had preceded the free-recall measure, these data cannot be directly compared with those of Exp. III, where free recall was the sole measure. The means for the three replications and overall mean scores are shown in Table 1. Again, the abstract-concrete dimension was shown by analysis of variance to be significant, $F\ (1, 27) = 14.7$, $p < .01$, while the main effect of frequency showed no significance. Two interactions were found to be significant; Abstractness-Concreteness × Frequency, $F\ (1, 27) = 6.6$; $p < .05$, and the triple interaction, $F\ (2, 27) = 7.4$, $p < .01$.

DISCUSSION

Two findings from these experiments seem worthy of further note. First, the degree of abstractness of words, although a highly effective variable in paired-associate learning, meaningfulness measurement, and free recall, was shown to have no effect upon tachistoscopic recognition. Secondly, the frequency of usage variable, found to be highly effective in tachistoscopic recognition and paired-associate learning, was shown to have an unclear effect on measures of meaningfulness and on free recall. These contrasts are suggestive of differences in the behaviors involved in each of these situations.

The degree of abstractness of words, found by Gorman (1961) to affect short-term recognition recall, exerted a strong influence in the present study on paired-associate learning and free recall. Viewing differences in meaningfulness of abstract and concrete words as possibly basic to this finding, Exp. IV undertook to measure the relative meaningfulness of the four types of words used. That concrete words have a significantly greater number of associations than do abstract words suggests a greater availability of mediational links, enabling these words to be more easily associated with the nonsense syllables, as in Exp. II, and to be more easily tied together for later recall as in Exp. III and IV. In contrast with these effects was the absence of an effect of this variable on recognition thresholds. A similar lack of effect of meaningfulness upon recognition thresholds has been reported in two previous studies. Taylor (1958) found no effect on recognition thresholds of meaningfulness of nonsense syllables, whether defined in accordance with norms from the general population or defined by *S*'s own experiences in associating syllables with pictures of meaningful objects. Similarly Gibson et al. (1964) found meaningful nonsense syllables (such as IBM, NYU) to have higher recognition thresholds than meaningless but pronounceable syllables.

The variable of word frequency had a consistent effect on measurements of tachistoscopic recognition and paired-associate learning, with high-frequency words recognized more easily and learned more readily than low-frequency words. That frequency of usage is a potent variable for word recognition has been demonstrated in a number of previous studies (Howes & Solomon, 1951; McGinnies, Comer, & Lacey, 1952; Solomon & Postman, 1952; Taylor, Rosenfeldt, & Schulz, 1961). That high-frequency words should be more easily associated with nonsense-syllable stimulus items than low-frequency words seems consistent with, for example, experimental results reported by Underwood and Schulz (1960). Adopting the two-stage view of paired-associate learning (Underwood, Runquist, & Schulz, 1959), it might be of interest to determine whether the frequency effect acts upon the response-learning phase (high-frequency words having a greater probability of being encoded for future use) or upon the association phase (the greater meaningfulness of high-frequency words providing more mediating links). Evaluation of the relationship between frequency of usage and meaningfulness is relevant here. Cofer and Shevitz (1952) and Underwood and Schulz (1960) have presented evidence that a positive relationship exists. The finding of a significant positive correlation between frequency of concrete words and their meaningfulness, coupled with a zero correlation for abstract words, indicates that the frequency-meaningfulness relationship may vary as a function of other word characteristics. At the same time, the finding that the frequency-meaningfulness relationship covaries with the abstract-concrete dimension allows for no clear evaluation of the relevance of the frequency variable to either of the two stages of paired-associate learning.

The discrepancy between the strength of the frequency variable in paired-associate learning and its lack of a clear effect in free recall verifies Deese's (1961) prediction that "many of the relationships holding in free recall will not be found in paired-associates learning . . . [p. 27]." That frequency did not exert a strong effect in either of the two somewhat different free-recall measures is not surprising in the light of previous studies of this relationship. Peters (1936), using a narrow range of frequencies, found no effect of frequency on recall scores. Hall (1954), reasoning that if there is a relationship between perceptual thresholds and learning as measured by recall scores, then words of high frequency should be recalled better than words of low frequency, reported results indicating such to be the case. Similarly, Bousfield and Cohen (1955) and Murdock (1960) found high-frequency words to be recalled more easily than low-frequency words. On the other hand, Postman (1961) found no difference in short-term or long-term recall for high- and low-frequency words. Gorman (1961), measuring recognition memory for high- and low-frequency abstract and concrete nouns, reported high-frequency nouns to be less well remembered than low-frequency. Differences in method of measuring recall (Gorman, for example, used recognition), instructions (Gorman's *S*s were apparently unaware that they were to be tested), and other variables such as length of list make specification of the basis for obtained differences almost impossible. It is Deese's (1960) contention from his finding of a list-length-frequency interaction in free recall that ease of learning varies with interword associations—an important factor only with extremely high-frequency words and long lists. The relevance of interword associations to the free-recall measures in the present study could not be evaluated, since both Exp. III and Exp. IV measured recall for high- and low-frequency words in mixed lists, thus precluding a comparison of high- and low-frequency recall scores with the measure of interword association.

REFERENCES

BOUSFIELD, W. A., & COHEN, B. H. The occurrence of clustering in the recall of randomly arranged words of different frequencies of usage. *J. gen. Psychol.*, 1955, **52,** 83–95.

COFER, C. N., & SHEVITZ, R. Word-association as a function of word-frequency. *Amer. J. Psychol.*, 1952, **65,** 75–79.

DEESE, J. Frequency of usage and number of words in free recall: The role of association. *Psychol. Rep.*, 1960, **7,** 337–344.

DEESE, J. From the isolated verbal unit to connected discourse. In C. N. Cofer (Ed.), *Verbal learning and verbal behavior.* New York: McGraw-Hill, 1961. Pp. 11–31.

GIBSON, E. J., BISHOP, C. H., SCHIFF, W., & SMITH, J. Comparison of meaningfulness and pronunciability as grouping principles in the perception and retention of verbal materials. *J. exp. Psychol.*, 1964, **67,** 173–182.

GLAZE, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35,** 255–269.

GORMAN, A. M. Recognition memory for nouns as a function of abstractness and frequency. *J. exp. Psychol.*, 1961, **61,** 23–29.

HALL, J. F. Learning as a function of word frequency. *Amer. J. Psychol.*, 1954, **67,** 138–140.

HOWES, D. H., & SOLOMON, R. L. Visual duration threshold as a function of word probability. *J. exp. Psychol.*, 1951, **41,** 401–410.

MCGINNIES, E. M., COMER, P. B., & LACEY, O. L. Visual recognition thresholds as a function of word length and word frequency. *J. exp Psychol.*, 1952, **44,** 65–69.

MURDOCK, B. B., JR. The immediate retention of unrelated words. *J. exp. Psychol.*, 1960, **60,** 222–234.

NOBLE, C. E. An analysis of meaning. *Psychol. Rev.*, 1952, **59,** 421–430.

PETERS, H. N. The relationship between familiarity of words and their memory value. *Amer. J. Psychol.*, 1936, **48,** 572–584.

POSTMAN, L. Extra-experimental interference and the retention of words. *J. exp. Psychol.*, 1961, **61,** 97–110.

SOLOMON, R. L., & POSTMAN, L. Frequency of usage as a determinant of recognition thresholds for words. *J. exp. Psychol.*, 1952, **43,** 195–201.

SPIELBERGER, C. D., & DENNY, J. P. Visual recognition thresholds as a function of verbal ability and word frequency. *J. exp. Psychol.*, 1963, **65,** 597–602.

STOKE, S. M. Memory for onomatopes. *J. genet. Psychol.*, 1929, **36,** 594–596.

TAYLOR, J. A. Meaning, frequency and visual duration thresholds. *J. exp. Psychol.*, 1958, **55,** 329–333.

TAYLOR, J. A., ROSENFELDT, D. C., & SCHULZ, R. W. The relationship between word frequency and perceptibility with a forced-choice technique. *J. abnorm. soc. Psychol.*, 1961, **62,** 491–496.

UNDERWOOD, B. J., RUNQUIST, W. N., & SCHULZ, R. W. Response learning in paired-associate lists as a function of intralist similarity. *J. exp. Psychol.*, 1959, **58,** 70–78.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning.* Chicago: Lippincott, 1960.

(Received April 27, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 2, 169–175

FUNCTION AND PERCENTAGE OF OCCURRENCE OF RESPONSE MEMBERS IN PAIRED-ASSOCIATE LEARNING

ROBERT K. YOUNG AND CARL I. FUHRMANN

University of Texas

4 groups of 16 Ss each served in a paired-adjective learning experiment which combined factorially the variables of occurrence of response members (100% or 40%) and function (single or double). Significant differences in rate of learning were found as a function of these 2 variables as well as the interaction between the 2. The 40% double-function condition (in which each item served as a stimulus in one pair and as a response in another) was found to be much more difficult than the other 3 conditions. Stage analyses indicated that considerable interference occurred in the double-function conditions after S-R associations had been formed. The results of the experiment were explained in terms of competing S-R and R-S associations.

The learning of single-function (SF) and double-function (DF) paired-associate lists has been compared in several studies (Primoff, 1938; Umemoto & Hilgard, 1961; Young, 1961). A DF paired-associate (PA) list is defined as one in which each item serves both as a stimulus in one pair and as a response in another. An SF, PA list is the more common one in which each item serves as a stimulus or response but not both. The research cited above has consistently found that learning a DF list is more difficult than learning an SF list.

Explanations of the difficulty of DF learning have been advanced by each of the authors cited above. In general, the authors agree that through S-R and R-S, or "backward" learning two different and competing associations are formed to each item in the list. The first association, symbolized B-C, is the S-R association commonly assumed to be formed during PA learning. The second association, symbolized B-A, is the R-S association formed during the learning of an A-B pair which interfers with the correct B-C association. Before the criterion of learning can be reached, the relative strength of the competing and incorrect B-A associations must be reduced through differential reinforcement. An experiment by Young and Jennings (1964) testing this competition-of-associations hypothesis found that transfer was negative when measured from a DF list to a list which was composed of the same items but which was "flipped" over to form an R-S list. Relatively strong, and in the second list incorrect, B-C associations and relatively weak, and correct, B-A associations would be expected to produce the negative transfer observed by Young and Jennings.

Within recent years, several studies using SF lists, e.g., Schulz and Runquist, (1960); Goss, Nodine, Gregory, Taub, and Kennedy (1962), have investigated the influence of percentage of occurrence of response members (% ORM) in PA learning. In these experiments the stimulus is presented and then the correct response is presented a predetermined percentage of the time. While difficulty of learning has been generally found to

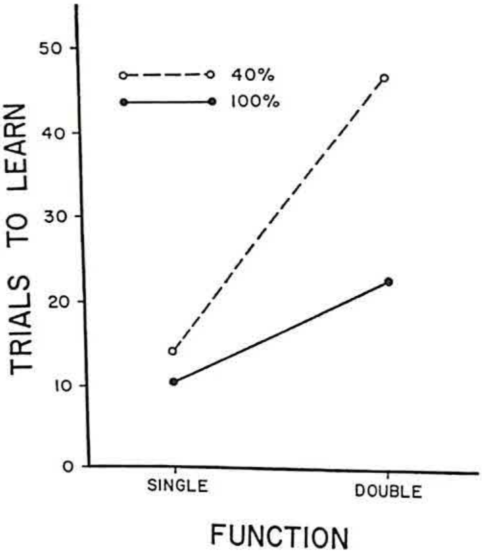


FIG. 1. Mean trials to learn in each of the four conditions.

be inversely related to % ORM, differences in rate of learning between even the highest and lowest percentage of conditions have been found to be relatively small.

The variables of % ORM and function, single and double, have been included in the design of the present experiment. A prediction about the influence of % ORM in DF learning, based on previous % ORM results, could be made in terms of differential response familiarization. From this point of view the differences between % ORM conditions obtained in SF learning should decrease in DF learning. The reason for this is that reresponse familiarization is directly related to % ORM in SF learning but the effect of this should be reduced considerably in DF learning since each response item appears every trial as a stimulus.

METHOD

Design.—Two % ORM conditions (100% and 40%) and two function conditions (SF and DF) were combined factorially. Within each of these four conditions, two lists were constructed from each of two samples of items, yielding a $2 \times 2 \times 2 \times 2$ factorial design. Each experimentally naive S learned a single list. A total of 64 Ss ($N = 4$ per group) from introductory psychology classes served in the experiment and were assigned to conditions in order of appearance according to a predetermined random sequence.

Materials and procedure.—Two sets of 10 adjectives were taken from Melton and Safier's list (Hilgard, 1951). Similarity, both meaningful and formal, among all 20 items was as low as careful inspection would permit. Two SF lists of five pairs each were constructed from each set of 10 adjectives with the restriction that the same five items serve as responses in the two lists within a set. From the five responses in each set two DF lists, also of five pairs each, were constructed.

Each list was presented in 10 different orders to prevent serial learning. In the case of the 40% groups, two responses occurred per trial, and each response occurred twice within the two blocks of five trials. In the case of the DF lists, pairs containing the same items did not follow one another.

Standard instructions for PA learning were given. In addition, Ss in the 40% groups were told that the response items would not always appear. All lists were presented on a Lafayette memory drum at a 2:2-sec. rate with a 4-sec. intertrial interval and were learned to a criterion of one perfect recitation. The Ss who failed to reach this criterion in 60 trials were arbitrarily given a score of 61. Two Ss were dropped for failure to follow instructions.

RESULTS

Learning.—The mean numbers of trials to learn the lists in each of the four conditions are presented in Fig. 1. The groups were found to have heterogeneous variances ($F_{\max} = 352.21$, $k = 16$, $df = 3$, $p < .05$) but the data were not transformed. In the analysis of the data it was found that lists in the 100% conditions were learned faster than those in the 40% conditions, $F(1,48) = 20.18$, $p < .001$, and that lists in the SF conditions were learned faster than those in the DF conditions, $F\ (1,\ 48) = 52.58$, $p < .001$. These results agree with the previously cited research. Finally, it was found that the interaction between the % ORM and function variables was significant, F (1, 48) = 11.09, $p < .01$. No other Fs were found to reach significance.

The apparent source of the significant interaction was the extreme difficulty of the lists in the 40% DF condition. The difficulty of these lists is also indicated by the fact that six of the seven *S*s who were unable to learn in 60 trials came from this condition.

Stage analyses.—Underwood, Runquist, and Schulz (1959) have suggested that list learning may be divided into two successive stages. The first, a response-learning stage, is measured by the number of trials before the response is first given. The second, an associative or "hook-up" stage, is measured by the difference between the trials on which the response is first given and first given correctly.

Analyses of these two stages were performed to provide additional information about the extreme difficulty of the 40% DF condition. These data are presented in Fig. 2. In the analysis of the response-learning stage, there was no difference between SF and DF conditions, F (1, 48) = 2.48. There were differences as a function of % ORM, with responses being given sooner in the 100% conditions than in the 40% conditions, F (1, 48) = 9.83, $p < .01$. In addition, the interaction between % ORM and function was significant, F (1, 48) = 4.69, $p < .05$. The significance of the interaction is attributable to the 100% DF condition in which responses were given earlier in learning than in the other three conditions. One obvious explanation of this phenomenon is that there were fewer items to learn in the DF lists than in the SF lists.

The second analysis was concerned with the associative stage. In contrast to the first analysis, the % ORM variable was not significant ($F = 2.63$) while the function variable was, F (1, 48) = 10.94, $p < .01$. The associative stage took longer in the DF than in the SF conditions. In addition, the interaction between function and % ORM was not significant ($F < 1.00$).

Neither stage analysis reflects with any degree of accuracy the relative difficulty of the lists. We would then conclude that much of the list difficulty occurred after the associative stage, as measured here, ended. For this reason an analysis of a third stage, a response-consolidation stage, was performed. The consolidation stage was defined as the difference between the trial on which the response was first given correctly and the trial

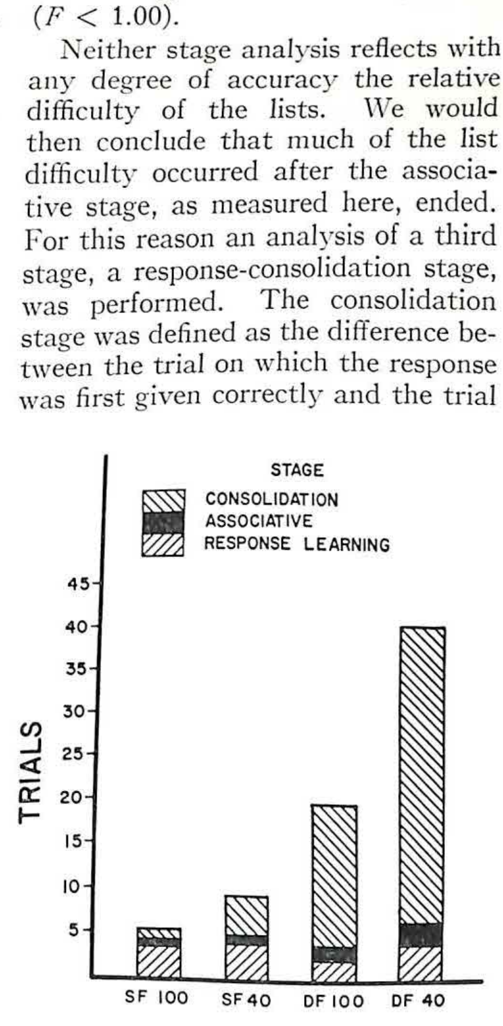


FIG. 2. Mean trials to reach the end of the response-learning stage, the associative stage, and the response-consolidation stage in each of the four conditions.

on which it was last given incorrectly. This is, of course, similar to the oscillatory cycle described by Hull (McGeoch & Irion, 1952, p. 117).

The data of the third analysis are also presented in Fig. 2. The results of this analysis indicate that function, % ORM, and the interaction between the two were significant, F (1, 48) $= 61.07$, 14.25, and 9.17, $p < .01$, respectively. The difference between the function conditions indicates that the consolidation stage took longer in the DF conditions than in the SF conditions. In addition, the analysis indicates that the 40% conditions yielded a longer consolidation stage than did the 100% conditions. Finally, as was the case in the trials-to-learn data, the significant interaction indicates that the consolidation stage takes longer in the 40% DF condition than would be expected from knowledge of the other three groups.

Both analyses of the associative and response-consolidation stages allow for a possible bias resulting from the fact that there were fewer DF items than SF items. In the DF conditions *S*s would have a higher probability of getting items correct by guessing than would *S*s in the SF conditions. This should result in a reduction of the length of the associative stage and an increase in the length of the response-consolidation stage. Since the associative stage took longer in the DF than in the SF conditions, no reanalysis of the second stage is necessary. On the other hand, the response-consolidation stage took longer in the DF than in the SF conditions as would be predicted by the guessing bias. To control for this, the data were reanalyzed. The *only* difference between this and the first analysis of the third stage is that the consolidation stage in the DF conditions is defined as the difference between the *second* correct response and the last incorrect response. The results of this analysis are nearly identical to the first with function, % ORM, and the interaction between the two being significant, F (1, 48) $= 46.20$, 13.26, and 8.27, $p < .01$, respectively. It does not appear that the results of the analysis of the response-consolidation stage are attributable to a bias in the data.

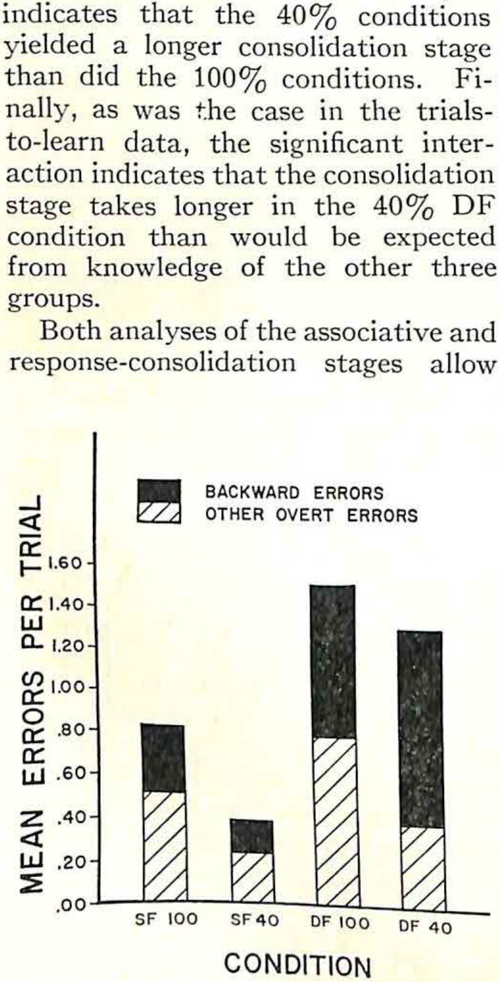


FIG. 3. Mean errors per trial in each of the four conditions. (See text for a more complete description.)

Errors.—The mean overt errors per trial for the four conditions are presented in Fig. 3. It can be seen that more errors per trial were made in the DF than in the SF conditions, with $F\ (1, 48) = 54.01$, $p < .001$, and that more errors per trial were made in the 100% than in the 40% conditions, with $F\ (1, 48) = 8.13$, $p < .01$. This latter finding is in agreement with previous research (Schulz & Runquist, 1960). No other F approached significance.

An attempt was made to determine the relation which errors had to other differences observed among the four conditions. The backward errors (incorrectly elicited B-A associations) per trial in the DF conditions were analyzed and no difference was found between the 100% and 40% conditions ($F = 1.61$). As a control, those errors in the SF conditions which *would have been* backward errors in the DF conditions were computed. More backward errors per trial were given in the DF conditions than in the SF conditions, with $F\ (1,\ 48) = 58.21$, $p < .001$. As shown by a median test, the proportion of backward errors to total overt errors was greater in the DF conditions than in the SF conditions, $\chi^2(1) = 10.56$, $p < .01$.

Next, for each *S*, backward errors were subtracted from overt errors. This gives a measure of overt errors other than backward errors and is also illustrated in Fig. 3. The results of the analysis of "other overt errors" indicated that errors again were directly related to % ORM, $F\ (1,\ 48) = 17.83$, $p < .001$, and that more errors were made in the DF than in the SF conditions, $F\ (1, 48) = 5.92$, $p < .05$.

Discussion

The analysis of trials to learn replicates previous research in that both % ORM and function were found to be variables related to rate of PA learning. In addition, instead of reducing the difference between % ORM conditions as might be expected from a response-familiarization interpretation, the difference between % ORM conditions increased from SF to DF learning.

The hypothesis of competing B-C and B-A associations as proposed by Primoff (1938), Umemoto and Hilgard (1961), and Young (1961) to explain DF learning may be useful in interpreting the results of the present experiment. The interference hypothesis assumes that the strength of the B-C associations must be increased relative to the B-A associations before the learning criterion can be reached. When the B-C pair is presented, incorrect B-A associations would not be reinforced while correct B-C associations would be. Under the 40% DF conditions B is presented but is not followed by C on 60% of the trials. Since both A and C are strongly associated to B, there is little if anything in the learning situation which would argue for an increase in B-C strength relative to B-A on these trials. If differential reinforcement can occur in the DF conditions only when responses are presented, then the 100% conditions should be learned faster. A similar competition of associations does not occur in SF conditions, the difference between % ORM conditions should be less, and the observed interaction can be predicted from the hypothesis.

The stage analyses are also consistent with the competition-of-associations hypothesis. In the present experiment the analyses of the first two stages should show relatively small differences across conditions while larger differences should appear during the third stage when *S* is assumed to have formed competing associations. The SF conditions should have relatively little interference, and as a consequence only a few trials should be spent in response consolidation. On the other hand, the DF conditions should have a great deal of interference, especially the 40% condition in which extinction of interfering B-A associations is retarded, and response consolidation should take a large number of trials.

Fewer errors per trial are given in the 40% than in the 100% conditions, and these error data suggest that *S*s restrict their guessing behavior in the 40% conditions to those responses which have relatively high associative strengths. If so, this would account for the absence of a difference in the backward-error data of the DF groups since this type of error is assumed to represent a strong interfering association. In addition, the 100% groups gave more "other overt errors" than did the 40% groups. If these "other overt errors" do *not* represent strong incorrect associations, then these data also lend support to the notion that the 40% groups restrict their guessing behavior to relatively strong associations.

As was the case with the stage analyses, the error data are also consistent with the interference hypothesis. The difference between the SF and DF conditions appears to be largely attributable to backward errors occurring in the DF conditions. This, in turn, suggests that list difficulty in the DF conditions is attributable to interference between B-A and B-C associations in the response-consolidation stage.

Neither the analyses of the first two stages nor the error analyses gave any indication of the extreme difficulty of the 40% DF condition. This would not, however, be inconsistent with the hypothesis. Interference between B-A and B-C associations is assumed to begin with the formation of associations during the second stage and continue into the third stage. Thus, analyses of the first two stages should give relatively little indication of the difficulty of DF learning.

Although backward errors per trial were elicited with a high frequency in both DF conditions, the absence of a difference between the 100% and 40% conditions suggests that interference during the third stage was high in both conditions. Under the 100% condition, when a backward error was given, the correct response was always presented immediately thereafter and differential reinforcement could then occur. Such was not the case in the 40% condition, and failure to present the correct response might not only retard differential reinforcement but might also allow the incorrect B-A association, when elicited, to increase in associative strength.

That such an increase can occur when responses are not given is suggested by PA learning studies by Mattocks (Underwood & Schulz, 1960, pp. 272–278) and Postman (1964) in which *S*s were required to supply their own responses. When the correct response is not presented, incorrectly elicited B-A responses may similarly increase in response strength, but this cannot be determined from the available data. However, it is not crucial to the interpretation of the present experiment that the strengths of B-A associations increase, decrease, or remain the same. It is apparent that the DF conditions are extremely difficult. From the stage analyses it would appear that the source of difficulty in the DF lists is associated with the response-consolidation stage. From the error data it would appear that incorrect B-A associations formed by the start of the response-consolidation stage interfere with correct B-C associations. Finally, it would appear that failure to present the correct response in the 40% DF conditions retards differential reinforcement of B-A and B-C associations, possibly allowing an increase in the strength of the incorrect B-A association, and permits continuing competition between B-C and B-A associations.

REFERENCES

GOSS, A. E., NODINE, C. F., GREGORY, B. N., TAUB, H. A., & KENNEDY, K. E. Stimulus characteristics and percentage of occurrence of response members in paired-associate learning. *Psychol. Monogr.*, 1962, **76**(12, Whole No. 531).

HILGARD, E. R. Methods and procedures in the study of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 517–567.

MCGEOCH, J. A., & IRION, A. L. *The psychology of human learning*. (2nd Ed.), New York: Longmans, Green, 1952.

POSTMAN, L. Acquisition and retention of consistent associative responses. *J. exp. Psychol.*, 1964, **67**, 183–190.

PRIMOFF, E. Backward and forward associations as an organized act in serial and in paired associate learning. *J. Psychol.*, 1938, **5**, 375–395.

SCHULZ, R. W., & RUNQUIST, W. N. Learning and retention of paired adjectives as a function of occurrence of response members. *J. exp. Psychol.*, 1960, **59**, 409–413.

UMEMOTO, T., & HILGARD, E. R. Paired-associate learning as a function of similarity: Common stimulus and response items within the list. *J. exp. Psychol.*, 1961, **62**, 97–104.

UNDERWOOD, B. J., RUNQUIST, W. N., & SCHULZ, R. W. Response learning in paired-associate lists as a function of intralist similarity. *J. exp. Psychol.*, 1959, **58**, 70–78.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning.* Chicago: Lippincott, 1960.

YOUNG, R. K. Paired-associate learning when the same items occur as stimuli and responses. *J. exp. Psychol.*, 1961, **61**, 315–318.

YOUNG, R. K., & JENNINGS, P. C. Backward learning when the same items serve as stimuli and responses. *J. exp. Psychol.*, 1964, **68**, 64–70.

(Received April 27, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 2, 176–181

POSITIONAL CUES AS MEDIATORS IN DISCRIMINATION LEARNING[1]

SHELDON M. EBENHOLTZ

Connecticut College

32 Ss each learned 2 lists of 8 paired associates. Stimuli were either horizontal lines (L) varying in length or gray patches (G) varying in reflectance. Responses were 8 CVC syllables. Half of the Ss learned in the order GL, the remaining Ss in the order LG. Ss were further distributed into random (R) and same (S) conditions depending upon the relative positions of responses on each stimulus dimension. Under Condition S response positions were the same on initial and transfer tasks, i.e., the syllable paired with the lightest gray was also paired with the shortest line, etc. Syllables were approximately 3 positions displaced under Condition R. Learning of the transfer lists was significantly faster under Condition S. Results indicate that positional cues mediate discrimination learning. It was suggested that common processes determine serial-position effects of the present study and those typically obtained in serial learning.

Position learning may be defined as learning to respond to members of an ordered set of stimuli in terms of their location with respect to the beginning and end of the series. Position is assumed to be a property of an item strictly by reason of its membership in an ordered set and is not a characteristic of items independent of series membership.

Recent evidence (Ebenholtz, 1963a, 1963b; Winnick & Dornbush, 1963; Young, 1962) has shown that position learning plays an important role in the serial anticipation of items in constant temporal order. In principle, however, the effects of positional cues should not be restricted to items in temporal series, but should apply to all dimensions which permit ordering of the dimensional attribute as from least to most, beginning to end, etc. The present experiment tested the hypothesis that positional cues mediate discrimination learning with stimuli varying along one of two dimensions, i.e., reflectance and line length.

METHOD

Subjects and materials.—The *S*s were 32 undergraduates fulfilling a requirement of the introductory psychology course at Connecticut College. All were naive to the materials and purposes of the study. Each *S* learned in succession two eight-pair lists. One list contained stimuli made up of achromatic colors varying in reflectance from white to black. These were made up of $\frac{1}{2}$-in. diameter patches having Munsell values of 1, 3.5, 4.5, 5.5, 6.5, 7.5, 8, and 9. The second-list stimuli were horizontal lines drawn in black ink and having lengths of $\frac{1}{2}$, $\frac{7}{8}$, $1\frac{1}{2}$, $2\frac{1}{8}$, $2\frac{3}{4}$, $3\frac{3}{8}$, 4, and $4\frac{5}{8}$ in., respectively. The selection of magnitudes was essentially arbitrary, the intention being to produce a moderately easy task. Both the Munsell patches and line lengths were presented on 3×5 in. white index cards. A three-letter nonsense syllable was typed approximately 1 in. below the stimulus on each card. The syllables, taken from the Glaze (1928) list, varied in association value from 80 to 100% and were chosen to minimize response similarity. The same set of eight syllables was used with both sets of stimuli.

---

[1] This investigation was supported in part by Public Health Service Research Grant MH07114-01, from the National Institute of Mental Health.

The author is grateful to Jean Ebenholtz for her aid in the preparation of materials and to Barbara Brachman and Linda Foster for collecting the data.

Design.—Half of all *S*s started with the line-syllable pairs (Group LG) and the remaining *S*s started on the gray-syllable list (Group GL). The position-learning hypothesis requires that those responses which occupy corresponding positions on both stimulus continua should transfer more rapidly than responses which are displaced from their original locations. Accordingly, half of all *S*s in each group were assigned to either Random (R) or Same (S) conditions depending upon the relative placement of the positions of the response terms across the two lists. There were thus four groups of eight *S*s each. Assignment to each group was determined by the order of appearance of *S* at the laboratory and was made on an alternating basis in the order LGS, LGR, GLS, GLR, etc. Considering Groups LGS and GLS the relation between the initial and transfer lists may be represented as an A-B, A-B transfer paradigm providing the positions are regarded as the effective stimuli. By the same reasoning Groups LGR and GLR represent an interference paradigm, i.e., A-B, A-Br. On the other hand, if the stimulus is assumed to be specifically dependent upon nonserial characteristics, e.g., a given percentage of reflectance, the appropriate transfer paradigm would be A-B, C-B for all four groups, since line length and reflectance are dissimilar stimuli.

Eight different sets of stimulus-syllable pairings were formed so that each syllable appeared once as response member with each stimulus. There were thus eight different sets of eight gray-syllable combinations and eight sets of eight line syllable pairs. Syllables were in this manner counterbalanced across stimuli.

Groups with response terms at corresponding locations on both lists (i.e., Cond. S) were formed by selecting lists such that the syllable to be associated with the lightest gray would be paired with the shortest line length; the syllable paired with the next lightest gray was also paired with the next shortest line, etc. Under Cond. R transfer lists were chosen so that response terms were approximately an average of three positions removed from their original location on the first list. The net result of this procedure was that the groups as a whole learned the identical sets of line syllable and gray syllable pairs, but differed in the order of learning the lists (i.e., G to L or L to G) and/or the particular first list—second list combination (i.e., Cond. R or S).

The anticipation method was used throughout with a 3:3-presentation rate. Timing was by metronome and the cards were presented manually. The presentation sequence was systematically varied on each of eight trials and then repeated. In addition, each *S* within each group started with a different stimulus sequence. In *no* case was the stimulus sequence ordered from light to dark or small to large (or vice versa). Learning was to a criterion of eight correct responses on a single trial or 15 trials, whichever occurred first. Approximately 2 min. intervened between tasks.

RESULTS

Initial list.—Nonparametric analyses were undertaken because the criterion measures tended to be skewed. Mean errors per trial to discriminate the line stimuli averaged 3.98 and 4.00 for Groups LGS and LGR, respectively. Median trials to criterion were 14.5 and 14.0. In both cases $p > .05$ by the Wilcoxon test with unpaired replicates (Wilcoxon, 1949). Trials to criterion were also evaluated by the median test (Siegel, 1956) which yielded $p > .05$. Two *S*s in Group LGS and three in Group LGR failed to respond correctly to all eight items on a single trial within the 15-trial limit.

The *S*s discriminating between the various shades of gray in Groups GLS and GLR averaged 3.92 and 4.06 errors per trial, respectively. The median trials to criterion was 15.0 for both groups. Statistical analyses were made as indicated above and in no case did the groups differ significantly. Five *S*s in Group GLS and six in Group GLR failed to give all eight responses on a single trial.

Since *S*s contributing to Cond. S and R were found not to differ on first-list performance, it was possible to compare the relative ease of learning pairs containing lines and grays, respectively, as stimuli. Mean errors per trial for LGS and LGR combined were 3.99. This was also the com-

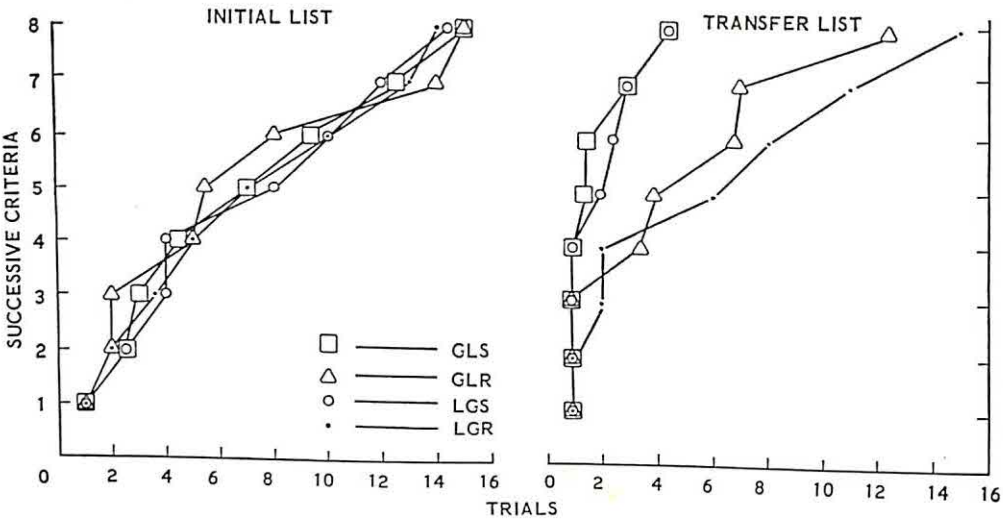


FIG. 1. Successive criteria as a function of median trials for initial and transfer lists under Cond. S and R.

bined average for GLS and GLR. The Wilcoxon test yielded $p > .05$. Trials to criterion were evaluated by the median test which was likewise not significant, $p > .05$. The number of *S*s failing to respond correctly to all eight stimuli on a single trial was 5 and 11 for lines and grays, respectively. There was thus some tendency for the former to be learned more rapidly.

Transfer list (*line stimuli*).—Mean errors per trial were 1.67 and 2.98 for Groups GLS and GLR, respectively. The Wilcoxon test yielded significance, $p < .02$. Median trials to criterion were 4.5 under Cond. S and 12.5 under Cond. R; by median test $p < .01$. All eight *S*s under Cond. S completed learning within 15 trials whereas three *S*s failed to complete learning under Cond. R.

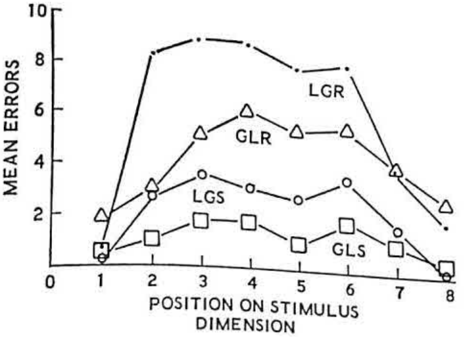


FIG. 2. Mean errors per position on transfer task for Groups LGR, GLR, LGS, and GLS.

Transfer list (*gray stimuli*).—Groups LGS and LGR produced means of 2.35 and 3.56 errors per trial. By the Wilcoxon test this was significant, $p < .05$. Median trials to criterion were 4.5 and 15.0 under Cond. S and R, respectively. The median test yielded significance, $p < .05$. One *S* failed to complete learning under Cond. *S* whereas five *S*s did not reach criterion under Cond. R.

Figure 1 shows the rate of learning of initial and transfer lists. Successive criteria were plotted as a function of the *median* trial at which each criterion was attained. A score of 15 was assigned wherever *S* failed to achieve a given criterion. Learning of the second list under Cond. S was decidedly more rapid than under Cond. R, with the greater differences occurring late in learning.

Serial-position effects.—In Fig. 2

are plotted mean errors as a function of position on the corresponding stimulus continuum. For this purpose errors corresponding to the shortest line and the lightest gray were represented at Position 1, those of the darkest gray and longest line at Position 8, etc. Transfer lists only are represented in Fig. 2, however bow-shaped curves also occurred in the error distributions of the initial lists.

The present design does not permit evaluation of absolute transfer; however, the groups can be compared in relative terms. Relative transfer was defined as the difference in performance (i.e., mean errors) between List 1 and List 2 expressed as a percentage of List 1. List 1 and List 2 refer to the *identical* lists but learned by different *S*s either as initial or transfer tasks, respectively. For example, transfer under Cond. S to the task involving line-length discrimination was based upon the differences in performance between Group LGS and Group GLS on List 1 and List 2, respectively, expressed as a percentage of the former. This procedure rests upon the assumption that Group GLS would have performed equivalently with Group LGS had each learned their respective lists in the same order. The same assumption holds for the relation between Groups LGR and GLR. Percentage of transfer over all eight positions combined was 81.55 and 33.80 for line stimuli under Cond. S and R, respectively. The corresponding values for transfer to the gray stimuli were 66.37 and 16.25. In each instance transfer was approximately 50% higher under Cond. S than Cond. R.

DISCUSSION

Conditions S and R differed only in the extent to which positional cues were available to mediate acquisition of the transfer task. The relatively rapid mastery of the transfer lists under Cond. S for both lines and grays, clearly supports the proposition that positional cues play a role in discrimination learning. This conclusion is consonant with the suggestion of Eriksen (1958) that "what *S* learns is not to recognize individual stimuli as such, but instead to establish a frame of reference for the series of stimuli as a whole [p. 358]." From the present view, in order to respond to position one must take account of the series as whole, particularly its boundaries.

The finding of positional cues in discrimination learning permits no inference regarding the role of absolute nonserial stimulus properties. For the present, therefore, discrimination learning with single-dimensioned stimulus material may best be analyzed by assuming a compound stimulus and by approaching the problem from the point of view of stimulus selection (Underwood, 1963).

Considering the present experiment as an instance of perceptual learning, it is in order to inquire as to whether specificity theory (Gibson & Gibson, 1955) or association theory (Postman, 1963) offers the more adequate interpretation of the present data. According to the former view, perceptual learning "consists of responding to variables of physical stimulation not previously responded to [Gibson & Gibson, 1955, p. 34]." If the positions of stimuli relative to their end points are assumed to represent just such a variable of stimulation, then it follows that specificity theory offers an appropriate account of discrimination learning in the present context. Postman (1963) has stated an alternative view which relies heavily upon the notion of response-produced cues. According to this view, "perceptual learning consists of the elaboration and differentiation of responses evoked by external stimulus patterns. Discrimination among a series of stimuli improves as the responses evoked by each member of the series become more specific and the response-produced cues more distinctive [p. 63]." In order to apply this interpretation to the present data it is necessary simply to identify position as a cue-producing response to which is associated the appropriate instrumental response, i.e., a nonsense syllable. It is necessary to add, however, that in order to account for the obtained differences in transfer the mediational response under question must also be assumed to possess all the ordinal and boundary characteristics of position specified at the outset. Otherwise it would be strictly fortuitous that two arbitrarily chosen sets of stimuli (e.g., lines and grays of Cond. S) come to elicit the identical sets of mediated responses.

It is clear that both association theory and specificity theory can encompass the present data. The two views differ essentially in terms of the locus of the position-learning process. For specificity theory it is a variable of proximal stimulation whereas association theory would equate position learning with the acquisition of special cue-producing responses having positional properties. Although translations into the terminology of association and specificity theory may be made readily, it may prove more fruitful to consider these approaches as most relevant to instances of perceptual learning in which the use of positional cues is kept minimal.

The occurrence of bow-shaped error distributions (e.g., Fig. 2) in the data of discrimination studies with single-dimensioned stimuli has been aptly documented by Murdock (1960). From the present view, his mathematical model is somewhat limited. First, although the curves derived from his model yield excellent approximations, they, without exception, attain the point of maximum error at the midpoint(s) of the series. This is not, however, universally obtained. Second, Murdock's formulation refers specifically to the distinctiveness of individual stimuli as the basis for the error distribution. The present data and that of a previous study (Ebenholtz, 1963a) suggest that it is the distinctiveness of *positions* which the stimuli occupy that is responsible for the bow-shaped curves and permits the deduction of transfer across dissimilar stimulus dimensions. This type of transfer cannot be inferred from the assumption that discrimination learning reflects the relative distinctiveness of individual items.

The interesting question concerning the relation between stimulus distinctiveness and discrimination of position remains to be investigated. The two processes are closely related in that if one cannot distinguish between two items in terms of their absolute properties then, in all likelihood, their respective locations within a series must likewise be confused. On the other hand, there is the possibility that items that may easily be distinguished on the basis of nonserial characteristics may fail to be discriminated within a series due to a difficulty in position discrimination.

The most general implication to be drawn from the present study is that the view is supported that position learning is responsible for bow-shaped error distributions obtained both in serial-learning and discrimination-learning studies. Indeed, available data suggest the general proposition that whenever *S* must discriminate on the basis of position along a given dimension, a bow-shaped curve will result.

REFERENCES

EBENHOLTZ, S. M. Position mediated transfer between serial learning and a spatial discrimination task. *J. exp. Psychol.*, 1963, **65**, 603–608. (a)

EBENHOLTZ, S. M. Serial learning: Position learning and sequential associations. *J. exp. Psychol.*, 1963, **66**, 353–362. (b)

ERIKSEN, C. W. Effects of practice with or without correction on discrimination learning. *Amer. J. Psychol.*, 1958, **71**, 350–358.

GIBSON, J. J., & GIBSON, E. J. Perceptual learning: Differentiation or enrichment? *Psychol. Rev.*, 1955, **62**, 32–41.

GLAZE, J. A. The association values of nonsense syllables. *J. genet. Psychol.*, 1928, **25**, 255–269.

Murdock, B. B., Jr. The distinctiveness of stimuli. *Psychol. Rev.*, 1960, **67**, 16–31.

Postman, L. Perception and learning. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 5. New York: McGraw-Hill, 1963. Pp. 30–113.

Siegel, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

Underwood, B. J. Stimulus selection in verbal learning. In C. N. Cofer & B. S. Musgrave (Eds.), *Verbal behavior and learning: Problems and processes.* Ne York: McGraw-Hill, 1963. Pp. 33–48.

Wilcoxon, F. *Some rapid approxima statistical procedures.* (Rev. ed.) Ne York: American Cyanamid, 1949.

Winnick, W. A., & Dornbush, R. L. Ro of positional cues in serial rote learnin *J. exp. Psychol.*, 1963, **66**, 419–421.

Young, R. K. Tests of three hypothese about the effective stimulus in serial learn ing. *J. exp. Psychol.*, 1962, **63**, 307–313.

(Received May 4, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 2, 182–188

RELATIONSHIP BETWEEN LATENCY AND REMOTENESS IN PREFERENCE JUDGMENTS[1]

MARSHALL G. GREENBERG[2] AND DAVID G. DOREN

University of Minnesota

The relationship between response latency and the remoteness of an individual's ideal point in preference judgments was studied. 30 college students made pair-comparison preference judgments of shades of gray from light, medium, and dark ideal points effectively manipulated by instruction. Some relationship was found between the order in which the judgment tasks were presented and the ideal points adopted. The major findings indicated that as an individual's ideal point becomes more remote from a stimulus pair, his response latency decreases while the response probability remains relatively constant. The relevance of this finding for models of preference behavior and similarities judgments is briefly discussed.

Recently Greenberg (1963) presented three models for preference behavior formulated within the framework of Coombs' (1964) unfolding technique. The unfolding technique conceives of both individuals and stimuli as represented by points on a unidimensional scale called a joint scale (J scale). It is assumed that when an individual is presented with a pair of stimulus alternatives and asked to state a preference, he compares their relative distances on the J scale from his ideal point and chooses the nearer alternative. Greenberg's models attempt to describe *how* these comparisons might be made and can be briefly summarized as follows: The *absolute difference model* states that the individual compares the two distances and the choice is mediated by their difference. The *ratio model* states that the choice is mediated by the ratio of the two distances, becoming more difficult as their ratio approaches 1. The *two-stage model* requires first that the individual decide whether the stimuli are unilateral (both on the same side of his ideal point) or bilateral (one on either side of his ideal point) to him. For a formal discussion of laterality see Coombs, Greenberg, and Zinnes (1961) or Coombs (1964). If the pair is a unilateral one, the individual simply decides which element of the pair possesses more or less of the attribute in question and, depending upon whether his ideal point is to the right or left of the pair, makes his choice accordingly. The model does not apply to bilateral pairs.

If one assumes that choice latency is a monotonically increasing function of difficulty, these models may be tested by experimentally varying *S*'s ideal point with respect to unilateral stimulus pairs. The absolute difference model implies no change in latency as the ideal point is moved away from a unilateral pair, while the ratio model implies an increase in latency, and the two-stage model a decrease (because the first stage is made easier, while the second stage remains constant in difficulty). An earlier pilot study employing only four *S*s tended to support the two-

---

[1] This research was supported in part by grants to the first author from the National Science Foundation (GB 2345) and the Graduate School of the University of Minnesota.

[2] Now at the Procter & Gamble Company, Cincinnati, Ohio.

stage model over the other two (Greenberg, 1963). This report describes a more comprehensive attempt to determine the relationship between response latency and the remoteness of the ideal point from a unilateral stimulus pair in a preference situation.

METHOD

Stimuli.—The stimuli were constructed from six Kodak flashed density filters, rated in equal intervals of .05 from .20 to .45,[3] and labeled for convenience A (lightest) through F (darkest). Quarter-inch circles were fitted into two holes punched $\frac{3}{8}$ in. apart in rectangles of 5 mil Mylar, which were then mounted in 2 × 2 in. slide mounts. Two complete decks of 30 slides were constructed, so that one deck could be rerandomized while the other was being presented.

Apparatus.—The stimuli were rear projected by a Sawyers 500R slide projector onto an 18 × 24 in. frosted glass panel set in a wall 3 ft. above the floor level. The *S* was seated in an adjoining room, approximately 30 in. from the screen. On a table directly in front of *S* and below the screen, a panel was mounted with two 1-in. push buttons. A small red warning light was also mounted in the panel between and above the push buttons. During the presentation of the stimuli, the following sequence of events took place: The warning light came on for 1 sec. With the cessation of the warning light, an Alphax shutter mounted directly in front of the projector opened and a slide was presented to *S*. In addition a "start" command was sent to a Beckman 5230P timer. With *S*'s response, the shutter closed, a "stop" command was sent to the timer, and the latency was printed out in hundredths of a second by a Beckman 1453 printer. The *S*'s response also advanced the next slide into position in the projector, and opened an auxiliary Alphax shutter mounted in front of a Viewlex projector which kept the screen illuminated between presentations. After a pause of about 2 sec., the cycle was repeated, except at the end of a deck of 30 slides, when there was a pause of about 5 sec. while *E* changed the slide cartridge. All the stimuli were presented slightly out of focus. This served the dual purpose of eliminating the rough edge around the patches of gray, and rendering invisible any small specks of dust that might accumulate on the slide surfaces.

Instructions.—The *S* was seated in front of the screen and was read the following instructions.

> (Task B—Stimulus comparison) On the screen in front of you will appear two circular patches of gray. You are to choose that patch which appears to be the darker of the two. If it is the one on the left, push the button on the left down firmly. If it is the one on the right, push the button on the right down firmly. Rest your hands on the panel like this, (demonstrate) and push the buttons with your forefingers. Try to reflect your true opinion as accurately as possible, but having made your decision, respond as quickly as you can. Speed and accuracy are of equal importance in performing this task. The red light in the center of the panel will come on as a warning signal 1 sec. before each pair of gray patches is presented. Remember, you are to choose that patch which appears to be the darker of the two. Any questions? Okay.

Each *S* was given a series of Task B trials, and was then read the instructions for the first of the three preference-judgment conditions, Light (L), Medium (M), Dark (D).

> (Task L, M, D—Preference) Okay, now your task will be somewhat different. This time you are to choose that patch of gray which *you prefer* as the better representative of a good light (medium, dark) gray. Try to reflect your true opinion as accurately as possible, but having made your decision, respond as quickly as you can. Remember, choose that patch which you prefer as the better representative of a good light (medium, dark) gray. Any questions? Okay.

Design and procedure.—Thirty *S*s were observed initially for two trials (60 presentations) under the Task B instructions. This procedure allowed *S* to "warm up" and become accustomed to the range of stimuli. Following this, each *S* was observed for five trials under each task, L, M, D. The order of presentation of instructions was completely counterbalanced, five *S*s being observed under each of the six possible permutations of the instructional conditions. The six groups so formed were labeled for convenience: I (LMD), II (DLM), III (MLD), IV (DML), V (LDM), VI (MDL). Five different permutations of the 30 stimuli were constructed, and

---

[3] Density $= 2 - \log_{10}\%$ T where % T is the percentage of light transmitted by the filter.

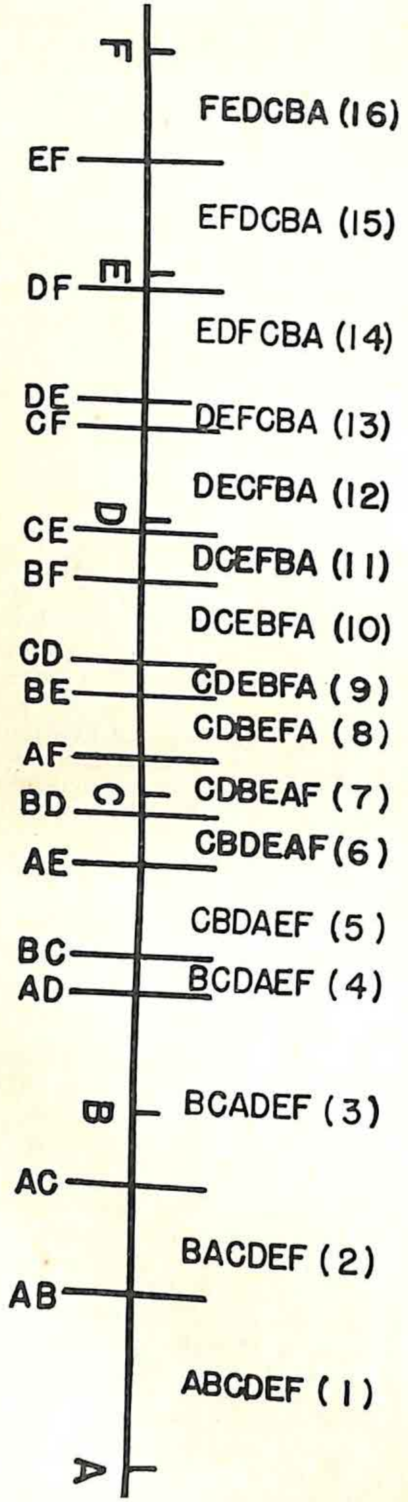


FIG. 1. J scale with all midpoints and I-scale intervals.

18 5 × 5 Latin squares were selected at random, 1 for each task within each of the six groups. Thus within each task, within each group, each *S* was presented the stimuli once in each of the five orders, and across the five *S*s the stimuli were presented once in each order on any given trial. In the analysis, observations on corresponding left-right pairs (e.g., AF and FA) were combined, yielding 10 observations for each pair for every *S* on all tasks.

Subjects.—Thirty-four undergraduate psychology students were given credit toward their final grade for participating in this experiment. Two *S*s were rejected for failure to follow instructions, and the data from two others could not be analyzed because of their many intransitivities. Because of the counterbalancing employed, the cells for these *S*s were immediately replicated, and the total kept at 30.

RESULTS AND DISCUSSION

The first stage in the treatment of the data required the determination of each individual's ranking of the stimuli in order of their nearness to his ideal point. This ranking represents his preference ordering and is called the individual scale (I scale). It is based upon the number of times a stimulus was chosen in a pair comparison over all replications within a given condition. The underlying J scale was assumed to be represented by the physical scale of percentage of light transmitted by the stimulus as shown in Fig. 1. The 15 midpoints between stimuli define a set of 16 I-scale intervals each of which has associated with it a unique preference ordering across the stimuli. These intervals are numbered for convenience of reference.

The I scales for each experimental condition are presented in Table 1. Of the 90 I scales obtained, it can be seen that only 3 cannot be made to fit the qualitative J scale and only 7 others fit the qualitative J scale but do not satisfy the metric on the scale. Thus most of the *S*s' ideal points can be readily located within a particular I-scale interval under each experimental task. As expected, most of the I scales which do not fit the quantitative J scale occurred in the Medium condition. This is because of the greater number of bilateral pairs and resultant greater degree of inconsistency on the part of *S*s (Coombs, 1958). Most of the *S*s chose one ideal point near each end of the scale and one in the vicinity of C or D near the middle. A notable exception is *S* IV5 who employed an ideal point near A in every task condition.

It is interesting to note the tendency for Groups I and V to choose ideal points in higher I-scale intervals under the Light condition than did the other groups. These two groups are the ones which received the Light condition first. It may be that the initial set of Task B trials tended to raise the adaptation level (see Helson, 1959) of *S*s in these groups, pulling their ideal points towards the darker

TABLE 1

I-Scale Intervals and Test Pairs for Each *S*

Group	*S*	Interval[a] (Light)	Interval[a,b,c] (Medium)	Interval[a,b,c] (Dark)	Test Pairs (Dark End)	Test Pairs (Light End)
I (LMD)	1	1	X	16	—	—
	2	4	4Q	16	—	—
	3	4	11	15.5	EF	AB
	4	5	8	16	EF	AB
	5	1.5	X	16	—	—
II (DLM)	1	1	8	15	EF, DE, DF	AB
	2	1	12	12	EF	—
	3	1	8	13	EF	AB
	4	1	4	16	EF, DE, DF	—
	5	2	10	8	—	—
III (MLD)	1	1	6.5	16	EF, DE, DF	AB
	2	1	5	16	EF, DE, DF	—
	3	1	8	16	EF	AB
	4	1	2	16	EF, DE, DF	—
	5	1	10	16	—	AB
IV (DML)	1	1	5	16	EF, DE, DF	AB
	2	1	1	16	—	—
	3	1	4	12	EF, DE, DF	—
	4	1	9	16	EF	AB
	5	1	1	1	—	—
V (LDM)	1	4	7Q	15	EF	AB
	2	3	4Q	X	—	—
	3	1	6Q	16	EF, DE, DF	AB
	4	3	8	14	EF	AB
	5	2	7Q	16	EF	AB
VI (MDL)	1	1	4	15	EF, DE, DF	—
	2	1	6	16	EF, DE, DF	—
	3	1	5	11	EF, DE, DF	—
	4	1	5	10Q	EF, DE, DF	—
	5	1	8Q	16	EF, DE, DF	AB

[a] A decimal entry indicates that due to a tie an I scale may be viewed as falling on a midpoint.

[b] Q indicates I scale fits qualitative J scale. Preceding number indicates closest interval approximation.

[c] X indicates I scale does not fit qualitative J scale.

TABLE 2

Mean Latencies in Milliseconds for Near and Remote Ideal Points on All Test Pairs

Group	*S*	AB			DE			DF			EF		
		M	D	(M-D)	M	L	(M-L)	M	L	(M-L)	M	L	(M-L)
I (LMD)	1	—	—	—	—	—	—	—	—	—	—	—	—
	2	—	—	—	—	—	—	—	—	—	—	—	—
	3	814	490	324	—	—	—	—	—	—	1080	779	301
	4	928	640	288	—	—	—	—	—	—	2534	2468	066
	5	—	—	—	—	—	—	—	—	—	—	—	—
II (DLM)	1	628	624	004	823	623	200	771	586	185	866	701	165
	2	—	—	—	—	—	—	—	—	—	1005	861	144
	3	694	811	−117	—	—	—	—	—	—	1168	1151	017
	4	—	—	—	655	719	−064	670	629	041	1036	958	078
	5	—	—	—	—	—	—	—	—	—	—	—	—
III (MLD)	1	846	459	387	967	669	298	979	636	343	1032	802	230
	2	—	—	—	806	648	158	774	509	265	894	756	138
	3	868	340	528	—	—	—	—	—	—	1115	641	474
	4	—	—	—	630	509	121	578	481	097	567	584	−017
	5	771	564	207	—	—	—	—	—	—	—	—	—
IV (DML)	1	929	507	422	927	942	−015	1171	661	510	1317	818	499
	2	—	—	—	—	—	—	—	—	—	—	—	—
	3	—	—	—	1011	993	018	1063	852	211	1537	1501	036
	4	663	549	114	—	—	—	—	—	—	859	741	118
	5	—	—	—	—	—	—	—	—	—	—	—	—
V (LDM)	1	714	570	144	—	—	—	—	—	—	797	723	074
	2	—	—	—	—	—	—	—	—	—	—	—	—
	3	528	479	049	552	663	−111	567	627	−060	665	730	−065
	4	615	523	092	—	—	—	—	—	—	729	1156	−427
	5	882	614	268	—	—	—	—	—	—	878	1013	−135
VI (MDL)	1	—	—	—	862	735	127	876	662	214	961	769	192
	2	—	—	—	700	615	085	646	562	084	728	777	−049
	3	—	—	—	863	734	129	848	700	148	1301	897	404
	4	—	—	—	870	620	250	977	570	407	1237	924	313
	5	739	714	025	732	561	171	894	649	245	735	578	157

end of the scale. This finding would be interesting to pursue in future research on preference behavior.

The final two columns of Table 1 list the critical stimulus pairs for each *S* from which data on the latencies of relatively near and remote ideal points were considered. These pairs were chosen as clearly unilateral to two ideal points which themselves were clearly distinct. Decisions as to which pairs could be considered unilateral were made by examining the fluctuations of each *S*'s ideal points across replications within a condition, and by subjectively determining his variability. Pairs were eliminated if either they appeared to be bilateral on some occasions, or if the two ideal-point distributions from which judgments would be compared appeared to overlap. Clearly some *S*s, because of the location or relative stability of their ideal points, provide more data relevant to the models under consideration than do others.

For the test pairs at the dark end of the scale the mean preference latencies from the Medium (nearer ideal) and Light (further ideal) conditions were compared and for the AB pair the latencies from the Medium and Dark conditions were compared. These latencies are presented in Table 2.

Of the 62 possible comparisons, 52 show a decrease and 10 show an increase in latency when the ideal point is more remote. A normal approximation to the binomial distribution yields a critical ratio of 5.2 for these data, $p < 10^{-6}$.

It appears quite clearly that, as an individual's ideal point becomes more remote from a pair of stimuli being judged, his response latency decreases. Assuming that this decrease in latency reflects a corresponding decrease in the difficulty of stating a preference, one is compelled to reject the absolute difference and ratio models in favor of the two-stage model.

One might, however, wish to consider an alternative interpretation of these data. It could be argued that the decreasing latencies result not from the decreasing difficulty of the decision, but from the lack of interest or concern in the discrimination when made from a remote ideal point. It might be that when an *S* is presented with two light grays such as A and B and is asked, "Which do you prefer as the better dark gray?" he responds quickly not because the discrimination is easy, but because he is "uninvolved" since neither stimulus is really a "good" dark gray. Such careless responding would certainly invalidate the results of this study. If this explanation were a valid one and *S*s were increasingly careless when making discriminations from remote ideal points, it would be reflected in the probabilities, which should reveal greater inconsistency with increasing remoteness. The data from this experiment are not as powerful as one might like to test this hypothesis for two reasons: (*a*) probability estimates are based on only 10 replications and are relatively unstable and (*b*) many of the proportions are 1.00 from both the near and remote ideal points, providing no discrimination at all, particularly for the AB, DE, and DF pairs which were easily discriminable. However, some information is made available by pooling the data across *S*s for each of the four test pairs. Among *S*s for whom AB was a test pair, the proportions of correct judgments from the near and remote ideal points were .986 and .993, respectively. For the DE pair the corresponding values were .900 and .923, for DF they were .938 and .954, and for EF, .727 and .759. None of the differences between these proportions approaches statistical significance and all are slightly in the direction of more rather than less consistent responses from the remote ideal points. These data clearly negate the hypothesis that *S* was careless in making his choices under the remote conditions.

This analysis of the probabilities, indicating relatively no change in the probability of a correct choice with increasing remoteness, is in agreement with a prediction of Luce's (1961) model for similarities judgments. One need only map the notion of ideal point from preference models into the reference stimulus in similarities judgments to see the isomorphism between these two types of judgment tasks.

While supporting the two-stage model to some degree, the present research points to the following problem. It would seem that both probability and latency should be monotonically related, although in opposite directions, to the psychological construct of discrimination difficulty. Yet in the present study, probability remains relatively constant while latency decreases with increasing remoteness of the reference or ideal point.

Luce's (1961) model employed probability as a measure of discrimination difficulty while Greenberg (1963) used latency to measure the same variable. Our data indicate that either (*a*) probability and latency do not measure the same type of difficulty or (*b*) latency, being more sensitive, reflects small differences in difficulty which probabilities are unable to detect. Further research of a parametric nature is needed to determine the exact nature of the relationship between latencies and probabilities in both preference and similarities judgments under a variety of experimental conditions.

REFERENCES

COOMBS, C. H. On the use of inconsistency of preferences in psychological measurement. *J. exp. Psychol.*, 1958, **55**, 1–7.

COOMBS, C. H. *A theory of data*. New York: Wiley, 1964.

COOMBS, C. H., GREENBERG, M. G., & ZINNES, J. A double law of comparative judgment for the analysis of preferential choice and similarities data. *Psychometrika*, 1961, **26**, 165–171.

GREENBERG, M. G. J scale models for preference behavior. *Psychometrika*, 1963, **28**, 265–271.

HELSON, H. Adaptation-level theory. In S. Koch (Ed.), *Psychology: A study of a science*. Vol. 1. New York: McGraw-Hill, 1959. Pp. 565–621.

LUCE, R. D. A choice theory analysis of similarity judgments. *Psychometrika*, 1961, **26**, 151–164.

(Received May 4, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 2, 189–191

VISUOMOTOR ADAPTATION TO DISCORDANT EXAFFERENT STIMULATION

I. P. HOWARD, B. CRASKE, AND W. B. TEMPLETON

University of Durham, England

Held and his co-workers claim that changes in the stimulation of receptors consequent upon self-produced movements (reafference) are essential for the development of visuomotor coordination. Stimulation of receptors in an inactive or passively moved animal (exafference) is claimed not to lead to the development of visuomotor skills. It is suggested that Held did not make any real effort to train his *S*s exafferently. In the present experiment, a rod seen through displacing mirrors was moved towards S until it hit him. This discrepancy between vision and touch (exafference) has been found to lead to some significant adaptation of active pointing towards the real position of targets seen through displacing mirrors.

Hein and Held (1962) have claimed that visuomotor adaptation to a displaced visual input can only occur under conditions where the self-produced movements of *S* are accompanied by a visual indication of the error of movement to a specified visual target. Such stimulation of exteroceptors related to the motor outflow involved in self-produced movement has been termed "reafference" by von Holst (1954). It is contrasted with "exafferent" stimulation which is that stimulation consequent only on events in the environment of *O*. Held and Hein (1958) found that passive movement of a seen hand which was displaced by prisms, did not lead to any improvement in the ability to point actively to that target. Active movement led to a rapid improvement.

Held's procedure provided very little exafferent information in the passive condition. Indeed, in one experiment (Held & Bossom, 1961) *S*s were merely wheeled about, and therefore were exposed only to a moving visual display seen through prisms. There was nothing to indicate to *S*s that their visual world was displaced. They could not even see their own bodies. Similarly the kittens which Held and Hein (1963) reared under passive conditions were given nothing to correlate with the moving visual display to which they were exposed.

In short, Held never really tried to teach his human or animal *S*s exafferently; so that his conclusion that exafferent stimulation is ineffective in the development of visuomotor coordination is unwarranted.

There are many ways in which exafferent stimulation may be used to indicate to an *S* the extent of any optical displacement introduced. For instance, *S* may be allowed to see bells which are ringing, or visible rods may come towards him and prod him, or he may be allowed a view of his own body, when he will "see" it in a different place to its "felt" position. Wallach, Kravitz, and Lindauer (1963) have in fact demonstrated that such passive training does lead to some visuomotor adaptation. Wooster (1923) also found that passive training led to adaptation, and a study of the conditions of her experiment indicates that visual asymmetry was probably the cause of this learning. Held and Bossom deliberately pre-

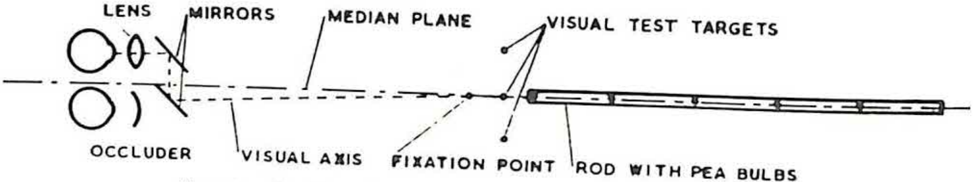


FIG. 1. Outline plan of the apparatus used in the study.

vented their *S*s from viewing their own body, without giving a good reason for this procedure. In any of these procedures, *S* must not be allowed to move actively.

The present experiment is designed to test whether prodding a person with a rod seen in a displaced position leads to visuomotor adaptation.

METHOD

Figure 1 shows the layout of the apparatus. The optical device consisted of two parallel mirrors which displaced the light from objects 2 in. to the left before it entered the left eye. Mirrors were used rather than prisms, because they do not introduce any apparent curvature, tilt, or color fringes. The displacement is parallel rather than angular, which is essential for our purpose; with prismatic angular displacement the apparent displacement reduces to zero as the viewed object comes towards *S*. Mirrors have one disadvantage: They lengthen the optical path and hence reduce the apparent size (or increase the apparent distance) of the visual target. This means that where a series of visual targets is used, as in our experiment, the apparent distance between the targets is distorted and visuomotor coordination will be disturbed accordingly. This disadvantage was overcome by introducing a lens system which magnified the displaced visual image to a size corresponding to its distance.

TABLE 1

MEAN ERROR IN INCHES OF POINTING AT TARGET OPTICALLY DISPLACED 2 IN. LATERALLY, BEFORE AND AFTER BEING TOUCHED AND NOT BEING TOUCHED

	Not Being Touched	Being Touched
Before	1.94	2.15
After	1.88	1.51
Adaptation	0.06	0.64

The rod consisted of a rigid wooden bar, the first 14 in. of which had five pea bulbs countersunk into the top surface at intervals of $3\frac{1}{2}$ in. The rod was mounted horizontally in the sagittal plane and could be moved towards the center of *S*'s lips. As seen through the mirrors, however, it appeared as if it would hit the face 2 in. to the left of the mouth. Nothing could be seen but the lights on the rod and a stationary fixation point. The lights used as visual targets in the test condition were three fine 1-in. high light slits displaced 2 in. apart in the frontal plane at a distance of 17 in. from *S*. The center light was in *S*'s objective median plane.

The *S*'s head was clamped in a headrest, the right eye was always occluded.

Each of the 20 *S*s (10 male, 10 female) were subjected to two conditions.

In Cond. I, *S* was asked to look through the optical device and point to the target light. These appeared one at a time in random order, 20 times in all. The *S* was allowed to fixate the slit displayed, and the hand could not be seen. This initial test established the pretraining deviation of *S*'s pointing in relation to the displaced visual targets. He was then told to remain still and to fixate a light just above the rod, 12 in. away. The rod was then moved from a distance of 14 in. until it hit *S* on the lips. This was repeated 20 times. The *S* was then immediately retested on the pointing task.

In Cond. II, the same procedure was employed except that in the training, the rod did not quite touch *S*. The order of conditions was alternated over *S*s with an interval of at least a 1 wk. between conditions.

RESULTS

The results are set out in Table 1. The "being-hit" training procedure produced a significant mean change in pointing of 0.64 ± 0.16 in. towards the actual position of the target lights, i.e., about one third of the displacement. This difference is significant, $t\ (19) = 3.9, p < .001$.

The "not-being-hit" condition produced no significant mean change in pointing. The size of the effect did not differ between sexes, nor between hands, though it tended to be larger for the right hand. Nor did the effect vary in size from the first to the last set of 10 judgments.

DISCUSSION

We have thus shown that discordant exafferent stimulation of an inactive *O* leads to some adaptation of pointing towards displaced visual targets. The effect could not have been due solely to the visually asymmetrical position of the rod, for this was present in Cond. II, where no adaptation occurred. The tactile stimulation was symmetrical, so that we had no need to have a control condition where *S*s were touched without being able to see the rod. Apparently all active movement was prevented, even convergence of the eyes. We were more careful in this respect than was Held in his "passive" conditions. We are forced to the conclusion therefore that discordant exafferent stimulation which gives an *S* "information" regarding optical distortion, may lead to at least some visuomotor adaptation.

REFERENCES

HEIN, A., & HELD, R. A neural model for labile sensori-motor co-ordinations. *Biological prototypes and synthetic systems.* Vol. 1. New York: Plenum Press, 1962.

HELD, R., & BOSSOM, J. Neonatal deprivation and adult rearrangement: Complementary techniques for analyzing plastic sensory-motor co-ordination. *J. comp. physiol. Psychol.*, 1961, **54**, 33–37.

HELD, R., & HEIN, A. Adaptation to disarranged hand-eye co-ordination contingent upon re-afferent stimulation. *Percept. mot. Skills*, 1958, **8**, 87–90.

HELD, R., & HEIN, A. Movement-produced stimulation in the development of visually guided behavior. *J. comp. physiol. Psychol.*, 1963, **56**, 872–876.

VON HOLST, E. Relations between the central nervous system and the peripheral organs. *Brit. J. anim. Behav.*, 1954, **2**, 89–94.

WALLACH, H., KRAVITZ, J. H., & LINDAUER, J. A passive condition for rapid adaptation to displaced visual direction. *Amer. J. Psychol.*, 1963, **76**, 568–578.

WOOSTER, M. Certain factors in the development of a new spatial co-ordination. *Psychol. Monogr.*, 1923, **32** (4, Whole No. 146).

(Received May 4, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 2, 192–200

DISCRIMINABILITY AND SCALING OF LINEAR EXTENT[1]

C. DOUGLAS CREELMAN

University of Toronto

Discrimination between lengths of printed lines was measured by 4 operationally different experimental techniques. 3 of the techniques (single stimuli, absolute judgment, and a modification of the method of constant stimuli) yielded psychophysical scales which were consistent with each other and with the theory of signal dectectability (TSD). In the 4th experimental procedure stimuli had unequal a priori probabilities. This produced consistently higher measures of discrimination. A possible explanation for this obtained discrepancy is the lack of long-term learning under this task as opposed to the others. The results extend the applicability of TSD to some further judgment tasks and experimental conditions.

The theory of signal detectability (TSD) has been shown to account for phenomena of simple visual discrimination (Swets, Tanner, & Birdsall, 1961), auditory signal detection (Licklider, 1959), speech discrimination (Egan, Clarke, & Carterette, 1956), and taste sensitivity (Linker, Moore, & Galanter, 1964). Where traditional methods of analysis showed sensory thresholds to change with experimental procedures, evidence has accumulated that detectability remains constant if one takes into account motivation and expectation, and their effects on the decision rules adopted by *O* which are independent of his sensitivity.

The task in the present experiments was judgment of line length. This is a qualitatively different task from those studied previously; it enabled easy manipulation to form different populations of alternative stimuli, while the stimuli were precisely specified in physical terms. Recently Weintraub and Hake (1962) reported results from a similar discrimination problem, and an analysis using the framework of TSD. Their stimuli were rectangles, and they studied the ability to tell a rectangle from a square. Pairs of alternative stimuli led to the same measures of *d′*, the measure of detectability, when different response scales were used, and when extraneous stimuli were introduced. The aim of the experiments below was to extend and to test the TSD framework in a wide range of decision tasks.

METHOD

The general features of the experiments will be described below, and the differentiating features of each procedure will be described separately.

Stimuli.—Horizontal lines about 3 mm. wide were printed on 3 × 5 in. white cards. A single line appeared on each card, centered in the vertical (3-in.) dimension. Lines were randomly varied in horizontal position so that *O* could not base his judgment of length on the relative position of the end point of the line.

Packs of 100 cards were made up for each experimental condition, and used throughout the experiment. The cards were placed after

---

[1] Most of this experiment was done under partial support of Contract No. NONR-248(55) between the office of Naval Research and the Johns Hopkins University. It was completed under Contract No. NONR-4010(03), and this is Report No. 4 under that contract. Reproduction in whole or part is permitted for any purpose of the United States Government.

The experiments were carried out while the author was on United States Public Health Service Postdoctoral Fellowship, No. MH 12, 140 at the Johns Hopkins University.

shuffling in a wooden holder, and one card at a time was presented by removing the top card of the pack, uncovering the next in sequence. There was no time limitation on the judgments; as soon as *O* had responded the result was tabulated on electrical counters at the same time that the correct response information was given and the next card uncovered.

The card holder was placed at the end of a long table, 70 in. from *O*. No attempt was made to fix *O*'s head or to restrict movement. Fluorescent diffused ceiling lighting gave uniform illumination over the experimental room of 25 footcandles (ftc.). The reflected illumination from the cards in the holder, which was at 70° with respect to the horizontal, was about 5 ftc.; the table was painted a medium gray color.

Observers.—Three male student *O*s worked individually 1 hr. a day, 4 days per week. They were paid an hourly rate, plus a bonus computed on the basis of performance. The *O*s were deliberately chosen for disparity in visual ability. One was quite myopic, and wore thick corrective lenses throughout the experiments, another was slightly astigmatic and wore no correction, and the third had normal vision. Absolute performance level was not the main interest in these experiments. The *O*s of differing ability were used in an attempt to obtain generality of the relationships between different procedures, with the necessarily limited number of *O*s that could be handled.

At least four training sessions preceded collection of the data. The four experimental conditions were run concurrently. A pack of cards for one set of conditions was selected randomly, 10–20 warm-up trials were given, and then the pack was gone though twice, with a short rest period in the middle, before another experimental condition was begun. The order of conditions was different for each *O*.

PROCEDURES AND RESULTS

Method of single stimuli.—A pack of stimulus cards consisted of 50 lines of each of two lengths. The task was to state, for each presentation, whether the line was the longer or the shorter possible alternative.

At least 800 observations were collected for each pairwise discrimination for each *O*. The four-fold table showing the frequency with which each response was given to each stimulus was used to compute percentages, which were taken as conditional probability estimates. A normal-curve table was used to estimate scale distance from the probability that the longer line would be correctly called longer, and the probability of a "false alarm," that the shorter line would lead to a response of longer. Normalized deviation scores for the two conditional probabilities were added to give the discriminability between the two lines, d'.

Discrimination was measured between 6 cm. and lines longer by 1, 2, 3, and 4 mm. in separate card packs.[2] Likewise, packs of cards were made up for discrimination between 3 cm. and lines longer by 1, 2, and 3 mm. Not all the data could be collected, because in some instances performance was not discriminably different from perfect, or 100% correct. Here an additive procedure was used (Creelman, 1963) in which scale values are derived for larger stimulus differences by measuring successive differential discriminations and adding together the resultant scale values. Where this procedure was used, estimates of scale values based on at least two different differential measures were averaged where possible.

Detectability scales for the three *O*s are shown in Fig. 1; points shown with asterisks were the results of additive scaling. The lines fitted to the data are least-square fits with the rather stringent restrictions that the lines go through the origin, and that for each *O* the line for discrimina-

---

[2] Throughout the paper, for ease of discussion, these nominal line lengths will be used. This is not the actual length, which is given in Table 4. The discrepancy is mostly due to the printer's proclivity to measure things in *points*, a unit of measure equal to 1/72 of an inch. Lines were cut to the nearest ½ point.

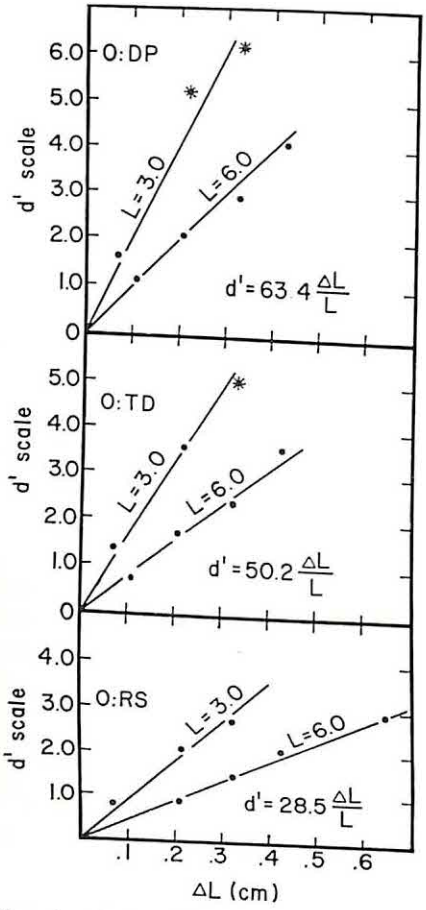


FIG. 1. Detectability scales for three *O*s for a difference of ΔL from 3.0 and 6.0 cm. (Lines represent a least-squares fit for each *O* with only one free parameter.)

tion from 3 cm. have twice the slope of that for discrimination from 6 cm. The resulting equations are shown on the graph for each *O*.

There are insufficient data here to conclude that Weber's law holds in general for linear discrimination. Many more base lengths than the two used would be needed. The form of the empirical equations provides a possibility for test, however. Note hat the equation $d' = K(\Delta L/L)$ is more general than the usual statement f Weber's law $\Delta I/I = C$, because he first statement predicts *performnce* rather than stimulus values which yield some fixed level of performance. Substituting $d' = 1.4$ into the performance equation (75% correct responses) we obtain values for the Weber fraction for the three *O*s of .022, .029, and .050, respectively.

Method of absolute judgment.—A pack of stimulus cards contained equal numbers of four possible line lengths, one with lines between 3.0 and 3.3 cm. and the other with lines 6.0 through 6.3 cm. The task was to specify on each trial which of the four possible line lengths was present, using the numerals 1–4. Each *O* went through each pack of cards at least 20 times, yielding 500 responses to each length. Feedback consisted of the correct response, which was given after each trial, and the data were tabulated in a 4 × 4 matrix showing the frequency with which each response had been given to each stimulus. These were treated by a procedure like that of Garner and Hake (1951). The stimulus-response, conditional-probability matrix was converted to equivalent standard scores, and the average scale distance between each pair of stimuli was calculated to yield scale values.

Table 1 shows the confusion matrices which were the basis for the scaling procedure. There are two matrices for each *O*, one for the shorter and another for the longer stimulus set. The table shows some unusual end effects. Usually the extreme stimuli of a set are identified correctly more often than those in the middle. The top confusion matrices to the left of Table 1, for DP and TD, show a *reverse* end effect. That is, *O*s correctly identified the middle two stimuli more often than they did the ends. Any interpretation in terms of anchors to which the judgments are referred will have difficulty with mixed data such as these. A more likely explanation is that *O*s, who had considerable sophistication after much practice in making judgments of line length, anticipated difficulty in discriminating the central members of the stimulus set, but overcompensated in setting their criteria. As we shall see, the differing criteria did not seem to change discriminability as measured by the scaling procedure. The proportion of correct responses is only part of the complete picture, and without the pattern of incorrect responses the question of whether discriminability is better at one place in the stimulus set relative to another is meaningless. The pattern of incorrect responses is taken into account in the detectability scaling analysis, the results of which will be presented along with the results of the third experimental procedure.

"Method of constant stimuli."—Quotation marks are used because there is no standard term for the procedure. The lines were of the same length as in the absolute-judgment problem, but the distribution was different and the response mode was changed. Within a pack of cards half the lines were "short," and the remainder were divided equally between the three longer lines. Only two responses were allowed in this case; that the line was short, or longer (without specifying which it might be). Feedback was the correct *response* for the trial, not specifying for a longer line, which one had

TABLE 1

CONFUSION MATRICES FOR ABSOLUTE-JUDGMENT DATA (PERCENT)

Stimulus	Response				Stimulus	Response			
					O:DP				
	3.0	3.1	3.2	3.3		6.0	6.1	6.2	6.3
3.0	73	27	0	0	6.0	66	31	3	0
3.1	14	80	6	0	6.1	21	46	30	3
3.2	0	7	83	10	6.2	4	26	53	17
3.3	0	0	25	75	6.3	0	4	31	65
					O:TD				
	3.0	3.1	3.2	3.3		6.0	6.1	6.2	6.3
3.0	46	49	4	1	6.0	45	43	10	2
3.1	11	64	24	0	6.1	16	48	30	6
3.2	1	11	69	19	6.2	4	24	53	19
3.3	0	2	42	56	6.3	1	8	40	51
					O:RS				
	3.0	3.1	3.2	3.3		6.0	6.2	6.4	6.6
3.0	57	31	10	2	6.0	50	35	14	1
3.1	26	42	28	4	6.2	21	36	37	6
3.2	4	21	45	30	6.4	2	18	49	31
3.3	1	5	36	58	6.6	1	4	33	62

TABLE 2

Confusion Matrices for "Constant-Stimuli" Data (Percent)

Stimulus	Response		Stimulus	Response	
	Short	Long		Short	Long
O:DP					
3.0	91	9	6.0	88	12
3.1	36	64	6.1	56	44
3.2	0	100	6.2	14	86
3.3	0	100	6.3	3	97
O:TD					
3.0	85	15	6.0	83	17
3.1	46	54	6.1	50	50
3.2	1	99	6.2	16	84
3.3	0	100	6.3	3	97
O:RS					
3.0	79	21	6.0	61	39
3.1	49	51	6.1	51	49
3.2	10	90	6.2	33	67
3.3	1	99	6.3	13	87

been present. Thus on half the trials the same length of shorter line was presented, and on the other half of the trials one of three longer lines appeared. (The proportion was actually 50–51, of course.)

The analogy with the more traditional method of constant stimuli should be clear. The major difference is that in the present case fully half the trials were "catch trials" while different sized increments were added for the remainder.

The data were tabulated in a 4 (stimuli) × 2 (responses) matrix, the conditional proportion of each response, given each stimulus. These data are shown in Table 2 for the three *O*s. A scale was constructed using the same techniques as with the absolute-judgment data; scale values could not be computed where the longer lines were correctly identified nearly 100% of the time.

The results of the three scaling procedures are summarized in Fig. 2, where scale values are plotted for the absolute-judgment and constant-stimuli procedures. The lines were taken from Fig. 1, and were not fitted to the new data. We can conclude tentatively, on the basis of Fig. 2, that the three judgment tasks are equivalent when detectability analysis is applied. A further task did not yield data consistent with the detectability framework, however.

Unequal stimulus probabilities.—Packs of cards were made up with

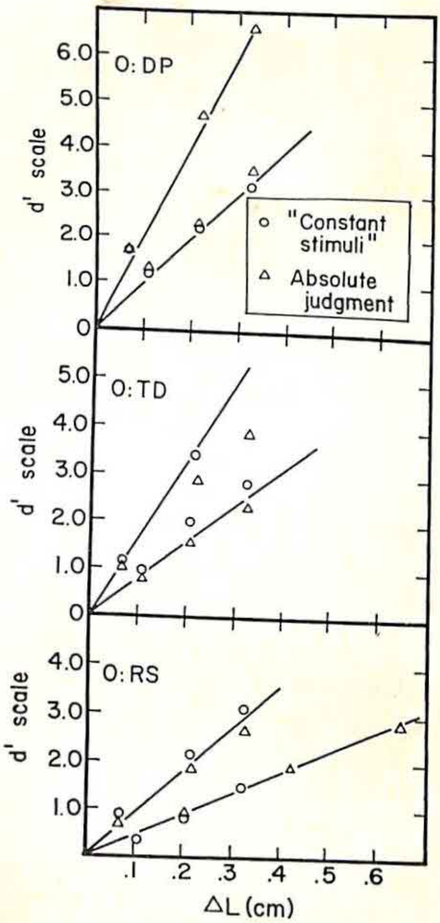


Fig. 2. Detectability scales traced from Fig. 1, with data from absolute-judgment and constant-stimuli procedures.

only two possible lengths, but the relative number of alternatives was changed. Longer and shorter lines were present in different packs with probability .75. The response was the same as in the single-stimuli task, to state for each card whether the line was the longer or the shorter. Again, feedback was given after each trial and a small bonus was paid for correct responses with equal fines for incorrect responses.

Four packs of cards were used; the longer or the shorter was the more frequent stimulus for each of two sets of lines: 6.0 vs. 6.2 cm., and 3.0 vs. 3.1 cm. At least 2,000 observations were collected for each of the four tasks for each *O*. The data were again in the form of four-fold conditional probability tables. Table 3 shows correct "long" responses and false alarm rates for each *O*. For comparison the table also shows proportions obtained with these line pairs in the equal-probability experiment, and resultant d' values.

In every instance the d' obtained with unequal stimulus probability was nearly the same whether the shorter or the longer was the more likely alternative. The values of d' were larger than those obtained with equal probabilities for five of the six comparisons. An analysis of variance was performed, using the scale values from these two discriminations (3.0 vs. 3.1, and 6.0 vs. 6.2), over all four experimental conditions, and it showed *O*s and experimental conditions to be the only significant sources of variance. The unequal-probability condition was the only experimental condition that was significantly different from the others.

TABLE 3

RESULTS FROM UNEQUAL PROBABILITY OF ALTERNATIVES

	3.0 vs. 3.1			6.0 vs. 6.2		
	$P_L(L)$	$P_S(S)$	d'	$P_L(L)$	$P_S(S)$	d'
O:DP						
50–50	.81	.76	1.62	.85	.84	2.06
25–75	.93	.69	1.98	.95	.80	2.49
75–25	.66	.93	1.92	.77	.97	2.62
O:TD						
50–50	.74	.77	1.38	.79	.81	1.68
25–75	.91	.49	1.33	.93	.69	1.98
75–25	.52	.89	1.29	.68	.93	1.99
O:RS (3.0 vs. 3.2)						
50–50	.83	.86	2.01	.70	.63	.83
75–25	.96	.79	2.52	.89	.43	1.06
25–75	.77	.96	2.55	.41	.91	1.08

DISCUSSION

Another way to test the significance of the difference in performance with the last experimental condition is to construct receiver-operating characteristics (ROC curves) from the data. The ROC curve plots the probability of correctly identifying the presence of a signal as a function of the probability of incorrectly stating that a signal was present when in fact it was not (false alarms). Here coordinates were the probability of correctly identifying a long line $P_L(L)$, and of responding long when the stimulus was in fact short $P_S(L)$. Figure 3 shows two representative ROC curves, one for discrimination between 6.0 vs. 6.2 cm., and the other for discrimination between 3.0 vs. 3.1 cm. There are three kinds of points for each set of data. The round points are from Table 3. Points near the negative diagonal show performance when the two alternatives were equally likely; points to the right and at the top show performance when the longer was the more probable alternative, and the points to the left at the bottom show performance when the longer was less probable. Lines of constant d' have been drawn through the central equal-probability points.

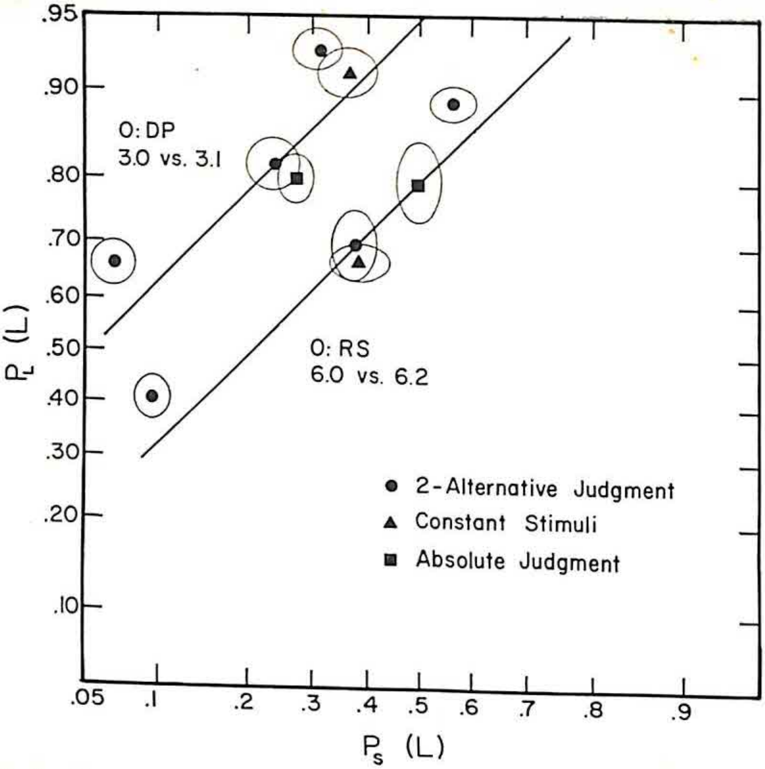


FIG. 3. Two representative ROC curves for discrimination between the same pair of stimuli under different experimental conditions. (Ellipses represent 2σ confidence intervals.)

The elipse around each point represents 2σ confidence intervals, on the assumption that binomial sampling variability applies to the obtained proportions. The assumption is most probably incorrect and the sampling distributions somewhat greater than the theoretical estimates. Nevertheless, the confidence intervals allow comparison of the data. The differences of the unequal-probability points from the lines of constant detectability are consistently large, and these differences are consistent across *O*s with the exception of TD, under one condition, as shown in Table 3.

Rectangular and diamond-shaped points in Fig. 3 represent performance with regard to the same pair of stimuli, from the other two experimental conditions. There was no consistent discrepancy here from the prediction of constant *d′*. The points were all near the line of constant *d′* as would be expected from the scales of Fig. 2.

Within the framework of TSD, changes in performance under different experimental conditions are considered to be the result of changes in *O*'s decision criterion. In order to yield a straight-line ROC curve, the underlying distributions must be normal. If they were heavily truncated we might expect performance with extreme criteria to be better than predictions on the basis of normal assumptions. However, the absolute-judgment task required multiple criteria, which in many cases were as far removed as those represented by the unequal-probability task, and here detectability did not appear to improve. There was no evidence which conclusively pointed toward nonnormality of the underlying distributions. Weintraub and Hake (1962) fitted their data with normal-normal ROC curves, and the results for three *O*s showed slopes near unity. Their *O*s used a rating response when the alternatives were two equally likely alternative stimuli. The discrimination task was sufficiently like that of the present experiment, and the results were close enough to those presented in Fig. 3, for the first three conditions, to raise serious question about the unequal-probability condition. It seems truly to be unique.

Learning to discriminate.—There is one possible explanation for the apparent greater discriminability in the unequal-probability tasks. Learning seems an important aspect of linear discrimination for all the conditions in the present experiments except this one. Long-term learning was not expected, and does not occur in other discrimination tasks using human *O*s (Swets & Sewall, 1963). This experiment was not designed to test or to isolate changes in performance with time; nevertheless, some rough comparisons were made. When relative performance under the different conditions was plotted as a function of the number of days *O* had been in the experiment, there was a distinct trend for performance to improve over the 45 experimental days in the first three tasks, but to remain the same for the unequal-stimulus probability task.

All conditions were not run on each experimental day and variability was large, with only 200 trials per day. Systematic comparisons could not be made. By the end of the experiment, however, performance was nearly the same for all the conditions. Observer TD, who showed the least improvement and the least difference in average performance, was a member of the college rifle team, and we might suppose that he had had considerable prior practice in making discriminations much like those required in the experiment.

Gundy (1961) found learning in an auditory-detection task when the signal was not specified for *O*s before each block of trials by producing the signal at a level considerably louder than that presented for detection. In the present experiments each run of 100 trials was preceded by 10 or more practice trials, but these were at the same level (the same lengths of lines) as were used subsequently. It may be that unequal probabilities lead to more precise specification of the more frequent stimulus, simply by dint of the greater amount of experience with that line length. A possible test of this notion might be consideration of sequential response probabilities (Atkinson, 1963). Unfortunately, the sequential data are not available to test this interpretation.

Measured line length, and performance of the ideal observer.—A sample of the printed stimuli was measured by vernier calipers, to the nearest .001 in. The results from measurement of 10 lines of each length are shown in Table 4. Along with each mean is an estimate of the population standard deviation; the lines were not all exactly the same length. Sources of variance probably were slight differences in the printing of the lines due to differing thicknesses of paper, pressure of the printing press, evenness of inking, or whatever. Bartlett's test for homogeneity of variance yielded an insignificant value of χ^2, between the .10 and .20 confidence levels, so the lines can be considered to have come from populations with the same variance. With this assurance the variances were averaged and the mean was used to compute an average standard deviation.

The final column in Table 4 shows the mean difference between each line length

TABLE 4

ACTUAL LENGTH OF STIMULUS LINES, AND ESTIMATED *SD*

Nominal Line Length	Actual Line Length (*M*)	*SD*	ΔL	d'
3.0	3.155	.0094	0	
3.1	3.226	.0085	.071	8.8
3.2	3.376	.0031	.221	27.4
3.3	3.477	.0072	.322	40.0
6.0	6.317	.0085	0	
6.1	6.426	.0075	.109	13.5
6.2	6.525	.0087	.208	25.8
6.3	6 645	.0109	.328	40.7
6.4	6.746	.0105	.429	53.2
6.6	6.970	.0070	.653	81.0

and the nominal 3 or 6 standard length, divided by the average *SD*. The ratio of such a physical mean difference divided by its standard deviation is a measure of the performance an ideally constructed *O* would show under the same conditions (Tanner & Birdsall, 1958). The performance equation for the ideal is $d' = 124\ L$. Observer efficiency, measured by the ratio between squares of observed and ideal performance ranged from $\eta = .03$ to $\eta = .000624$, for the best *O* under the easiest conditions to the poorest under the worst. Changes in observer efficiency with base length, *L*, cannot be attributed to differences in physical variance of the stimulus alternatives. The Weber law type of function reflects internal aspects of the discrimination process for judgments of linear extent.

Some limitations.—TSD clearly can be used for prediction of performance in a wide range of perceptual discrimination tasks. These range from the detection of signals in noise, to stimulus identification, to at least some kinds of loudness judgments (Creelman, 1963). In each case studied so far, the number of alternative stimuli was relatively small; pairs of signals were matched in loudness, one of no more than four alternative stimuli could appear in any single trial, and so forth. Garner and Creelman (1964) have suggested that when the number of alternative stimuli is large, factors independent of the discriminability between the alternative signals become critical in limiting performance. These factors have to do with maintenance of the alignment between response categories and stimulus alternatives, or stating it in another way, the maintenance of stable criteria. The point at which such response-criterion limiting factors become important is not known, although the number of alternatives at which difficulty could be expected will likely be near the number where information transmission breaks down on various perceptual continua (Miller, 1956).

REFERENCES

Atkinson, R. C. A variable sensitivity theory of signal detection. *Psychol. Rev.*, 1963, **70**, 91–106.

Creelman, C. D. Detection, discrimination, and the loudness of short tones. *J. Acoust. Soc. Amer.*, 1963, **35**, 1201–1205.

Egan, J. P., Clarke, F. R., & Carterette, E. C. On the transmission and confirmation of messages in noise. *J. Acoust. Soc. Amer.*, 1956, **28**, 536–550.

Garner, W. R., & Creelman, C. D. Effect of redundancy and duration on absolute judgments of visual stimuli. *J. exp. Psychol.*, 1964, **67**, 168–172.

Garner, W. R., & Hake, H. W. The amount of information in absolute judgments. *Psychol. Rev.*, 1951, **58**, 446–459.

Gundy, R. F. Auditory detection of an unspecified signal. *J. Acoust. Soc. Amer.*, 1961, **33**, 1008–1012.

Licklider, J. C. R. Three auditory theories. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 1. New York: McGraw-Hill, 1959. Pp. 42–75.

Linker, E., Moore, M. E., & Galanter, E. Taste thresholds, detection models, and disparate results. *J. exp. Psychol.*, 1964, **67**, 59–66.

Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.*, 1956, **63**, 81–97.

Swets, J. A., & Sewall, S. T. Invariance of signal detectability over stages of practice and levels of motivation. *J. exp. Psychol.*, 1963, **66**, 120–126.

Swets, J. A., Tanner, W. P., Jr., & Birdsall, T. G. Decision processes in perception. *Psychol. Rev.*, 1961, **68**, 301–340.

Tanner, W. P., Jr., & Birdsall, T. G. Definitions of d′ and η as psycho-physical measures. *J. Acoust. Soc. Amer.*, 1958, **30**, 922–928.

Weintraub, D. J., & Hake, H. W. Visual discrimination: An interpretation in terms of detectability theory. *J. Opt. Soc. Amer.*, 1962, **52**, 1179–1184.

(Received May 5, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 2, 201–209

REMINISCENCE AND FORGETTING IN A RUNWAY[1]

WINFRED F. HILL, ALBERT ERLEBACHER
Northwestern University
AND NORMAN E. SPEAR
Rutgers University

3 experiments attempted to explain the finding that runway speeds of rats declined within a day but increased from the last trial of 1 day to the 1st trial of the next. This reminiscence effect was found consistently with massed rewarded trials but not with spaced or with nonrewarded trials. Nonreward on a series of trials resulted in forgetting between days on a subsequent rewarded series, but reward magnitude on an initial series did not affect reminiscence on a subsequent series. Housing Ss in activity wheels had no effect on reminiscence.

Several experimenters (e.g., Hill & Spear, 1962; Jensen & Cotton, 1960; Renner, 1963) have noticed a tendency for runway speeds to decrease from trial to trial within a day and then increase from the last trial of 1 day to the first trial of the next day. This reminiscence effect (using "reminiscence" purely in an empirical sense to characterize a change in performance over a time interval), though familiar from human motor learning, is contrary to what would be expected from stimulus sampling theory (Estes, 1959) or from the concepts of warm-up and forgetting. The present series of experiments was undertaken in an attempt to explain this reminiscence.

Probably the most traditional possible interpretation would be in terms of reactive inhibition (Hull, 1943), which is assumed to build up during a series of massed trials and to dissipate between blocks of trials. This interpretation, however, faces at least two difficulties. (*a*) It assumes (using approximate values from Hill & Spear) that 2 sec. of running builds up more reactive inhibition than 20 sec. of rest can dissipate. (*b*) The concept of reactive inhibition has led to serious theoretical difficulties (Gleitman, Nachmias, & Neisser, 1954; Jensen, 1961). It therefore seemed worthwhile to look for some alternative explanation of the reminiscence effect.

EXPERIMENT I

It is possible that the reminiscence might be a result of activity drive (Hill, 1956). Perhaps rats kept in their fairly small living cages most of the time build up large amounts of activity drive, and on their first runway trial are motivated to run more by this drive than by hunger. If activity drive was rapidly satiated by this running, speeds would be expected to decline within days and recover between days, as observed. Experiment I tested this activity hypothesis.

Method

Subjects and apparatus.—The *S*s were 36 naive female albino rats of the Sprague-Dawley strain, 74–76 days old at the start of experimental training. The apparatus was the enclosed straight runway previously described by Lewis (1956) and employed by Hill and Spear (1962). It consisted of an 8-in. start box 6 in. wide, a 4-ft. alley 4 in.

---

[1] These experiments were conducted at Northwestern University and were supported by Grant GB-254 from the National Science Foundation.

wide, and a $10\frac{1}{4}$-in. goal box $9\frac{1}{4}$ in. wide. All parts of the apparatus were $8\frac{1}{2}$ in. high and painted black. A guillotine door 4 in. before the entrance to the goal box was used to prevent retracing. A 3 × 4 in. piece of sheet metal, painted black, was suspended by wire from the top of the goal box to mask the food cup from view as *S* approached. A Standard Electric timer was activated when the start-box door was raised and was stopped when *S* put weight on the floor of the goal box. The *S*s were housed in Wahmann activity wheels with small attached cages and in $9\frac{1}{2} \times 7 \times 7$ in. living cages, which have been described by Hill (1961).

Prehandling.—Each *S* received six daily 3-min. sessions of prehandling, the last one 48 hr. prior to the beginning of experimental training. During each session *S* was allowed to explore a large unpainted wooden box, presented with four of the 45-mg. Noyes pellets later to serve as reward, and picked up and replaced at least five times by *E*. Pellets uneaten were returned with *S* to the home cage along with a fifth pellet. A once-daily feeding schedule began on the first day of prehandling and was maintained throughout experimental training. The ration was 10 gm. of finely ground Purina lab chow and was presented 50–60 min. after the start of prehandling or experimental training.

Experimental training.—On each trial, after *S* was placed in the start box, the door was opened as soon as *S* was oriented toward it. The *S*s received 25 trials of acquisition and 10 trials of extinction, both 5 per day. During both acquisition and extinction, *S* was confined in the goal box for approximately 10 sec. following entry. Between trials on the same day, *S* was confined in its home cage, with water available, for 15–20 sec. One pellet in a food cup served as a reward. During extinction the food cup was removed from the goal box.

Differential treatment was introduced only with respect to the manner of housing during prehandling and between training sessions. A randomly determined 18 *S*s were housed in the activity wheels and the remaining 18 were housed in the living cages. All *S*s received acquisition training, but only 15 from each group received extinction.

Results

Speeds.—The acquisition data are shown in Fig. 1. The wheel-housed rats consistently ran faster (3.32 ft/sec) than the cage-housed (2.57 ft/sec), significant at the .01 level, *t* (34) = 2.81, for the last 5 acquisition

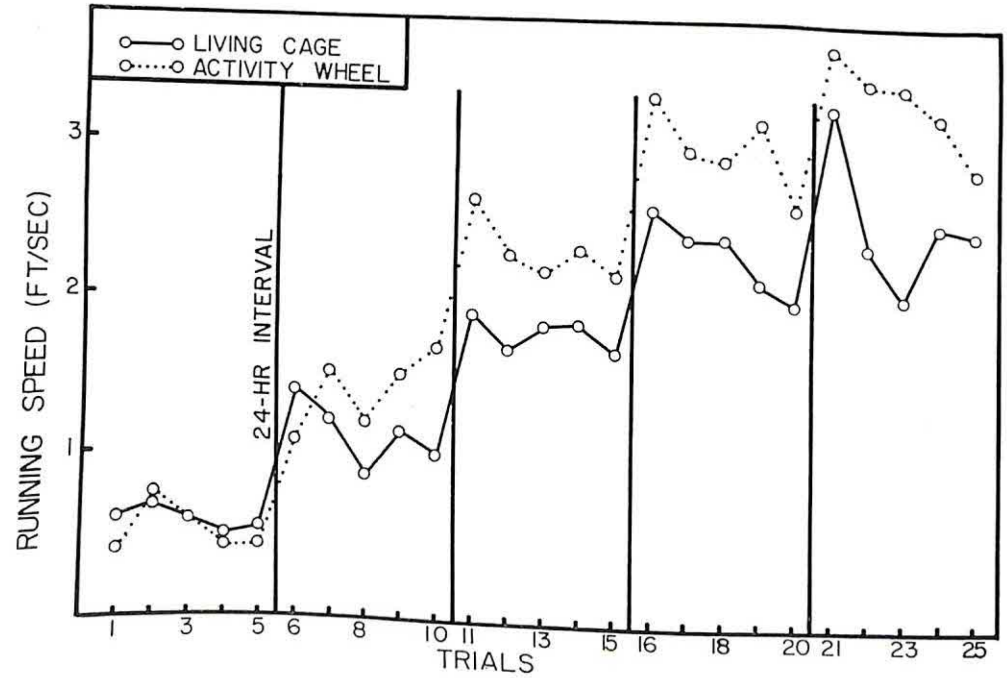


FIG. 1. Running speeds during acquisition in Exp. I (N = 18 per group).

trials. In extinction there was no difference in speed between the two groups.

Reminiscence.—The amount of (relative) reminiscence was computed as in Hill and Spear (1962). The *S*'s reminiscence score for a given day is obtained by subtracting the mean increase in speed from one trial to the next within the day (which is often negative) from the increase over the previous 24-hr. interval of no practice. (The working formula for obtaining this score is $\frac{1}{4}$ $(5F_c - 4L_p - L_c)$ where F_c, L_p, and L_c refer to speeds in ft/sec on, respectively, the first trial of the current day, the last trial of the previous day, and the last trial of the current day.) Thus a positive score indicates relative reminiscence and a negative score relative forgetting, while a zero score means that the 24-hr. interval has the same effect as the 20-sec. intertrial interval within a day. Reported scores are means averaged over days, with the first day necessarily omitted.

For both groups combined for Days 2–5, the reminiscence score was 1.03, significantly different from zero at the .001 level, $t\ (34) = 9.34$. However, a repeated-measures analysis revealed no significant effect of groups, days, or interaction. There was also significant reminiscence (which in this case is equivalent to spontaneous recovery) between the 2 days of extinction, mean = 2.56, $t\ (28) = 10.7$ for the combined group difference from zero, with no significant difference between groups.

To the extent that a learning curve is negatively accelerated, the measure of reminiscence would be slightly biased toward yielding positive scores. A formula completely free of either this or the reverse bias would be extremely complex. To test the importance of the bias, we recomputed the overall reminiscence, using a modification of the formula designed to produce a comparable bias in the opposite direction by subtracting the mean increase in speed within a day from the increase across the *subsequent* 24-hr. interval—$\frac{1}{4}(4F_S - 5L_C - F_C)$ where F_S refers to the first trial of the subsequent day. The effect of this change was negligible: The mean score for the first 4 days (the fifth necessarily being omitted) for the combined groups was 1.00, $t\ (34) = 9.21$, $p < .001$, for comparison with zero. The original formula was therefore retained in computing all subsequent reminiscence scores.

Discussion

The faster running by the wheel-housed rats is the opposite of what would be expected from differences in activity drive between the two groups. Although weights were not recorded, it seems likely from other data (Spear & Hill, 1962) that this finding is due to differential weight loss. The wheel-housed animals presumably lost so much weight from running in the wheels that their hunger drive was greater than that of the cage-housed *S*s, resulting in faster running in the alley.

The absence of any group difference in reminiscence argues against the hypothesis that activity drive is responsible for reminiscence. Although speed data are affected by the probable confounding of hunger drive and activity drive mentioned above, reminiscence scores should be unaffected by it. It appears, therefore, that the activity-drive explanation must be rejected. However, the highly positive mean reminiscence score confirms that this is a powerful phenomenon, still in need of explanation.

EXPERIMENT II

Hill and Spear (1962) noted that reminiscence was greater for *S*s receiving one-pellet reward than for those receiving four. This finding

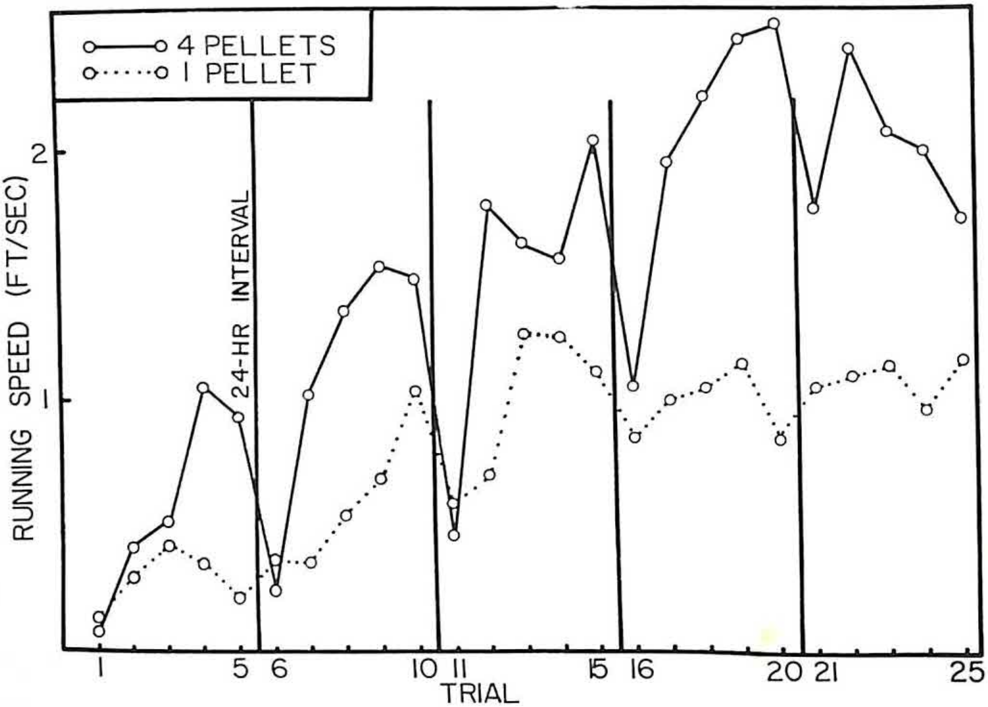


FIG. 2. Running speeds on the first 5 days of Stage 2 in Exp. II ($N = 7$ per group).

suggested the possibility that the effect might be even greater for rats receiving no food reward at all. To test this possibility, rats were given a long series of nonrewarded alley trials. In addition, to see what effect this experience would have on subsequent performance, they were then given ordinary acquisition training.

Method

Subjects and apparatus.—The *S*s were 15 naive female albino rats, 74 days old at the beginning of training. The runway and living cages were the same as in Exp. I, except that the shield was removed from the goal box, due to possible extraneous reinforcing effects resulting from the fact that *S*s sometimes jumped up and swung on it at the end of a run.

Procedure.—Pretraining was exactly the same as in Exp. I except that the daily ration was 9 gm., which continued throughout the experiment. In Stage 1, all *S*s received 75 trials at 5 trials a day. The first 13 days of training were successive, as were the last 2, but a period of 14 days intervened between these two blocks of training, during which deprivation scheduling continued but no training was given. No food or food cup was present in the goal box on any of these 75 trials. In Stage 2, all *S*s received 65 additional trials, 5 trials a day on successive days. A randomly selected seven of the *S*s received a reward of one 45-mg. Noyes pellet on each trial in this stage, seven others received four pellets, and the remaining *S* (scheduled to receive four pellets) was discarded for failure to eat the pellets. The trial-by-trial procedure was the same as in Exp. I. If an *S* failed to enter the goal box within 2 min., it was placed in the end box by *E*, and the time was recorded as 120 sec.

Results

Stage 1.—In the earlier part of the first stage, mean speeds showed fairly large cyclical fluctuations, but no clear pattern emerged. Later in this stage, mean speeds were uniformly low, partly reflecting the fact that there was a high proportion of nonruns (more than half of all opportunities on the last 15 trials). There was no evidence of reminiscence, the mean for Days 2–5 being only .05, t (14) from zero = .78.

Stage 2.—Speeds for the first 5 days of Stage 2 are shown in Fig. 2. As expected, the four-pellet group ran faster (2.00 ft/sec on the fifth day) than the one-pellet group—1.10 ft/sec, t (12) = 2.21, $p < .05$. Reminiscence, however, was not only absent but reversed; forgetting occurred between days. For consistency with Exp. I and with Hill and Spear (1962), reminiscence scores were analyzed separately for Days 2–5 (Fig. 2) and for Days 10–13 (not shown). In the earlier block of days, the means were −.83 for the four-pellet group and −.20 for the one-pellet group. The score was significantly negative for the four-pellet *S*s but not for the one-pellet *S*s, t (12) = 3.14 and .77, respectively. The difference in forgetting between the groups was not significant, t (12) = 1.67. In the later block of days, the above contrast between the groups was reversed. Both groups still had negative mean reminiscence scores, but now the one-pellet group was larger (−1.30, t = 5.49) than the four-pellet group (−.27, t = 1.15). In this case the group difference was significant (t = 3.07). At this point the four-pellet group was beginning to show a within-days decrement along with its forgetting, resulting in a suggestion of the slow first trial, fast second trial, and subsequent decline that is described by Logan (1960) as the typical within-days pattern. Though it is not at all typical in the present series of experiments, its occasional appearance may have theoretical importance. In comparing Logan's results with ours, it is noteworthy that the marked slowness of his *S*s on the first trial of a day did not appear on the early days. In addition, his *S*s were prehandled without reward, which may have a similar effect to giving initial nonrewarded "training" trials, as we did.

Discussion

The absence of reminiscence in Stage 1 indicates that the negative relation between reward magnitude and amount of reminiscence suggested by Hill and Spear (1962) does not include zero magnitude. This lack of reminiscence could be due either to the absence of reward or to the absence of a learning curve. Although neither reminiscence nor forgetting occurred during this operant-level period, the experience during this period had a great effect on the next stage, since forgetting appeared instead of reminiscence. This finding suggests that prior experience with reward or nonreward is an important determinant of short-term retention and one that should be more systematically studied.

EXPERIMENT III

The forgetting observed in Stage 2 of Exp. II does not seem to be a forgetting of the running response itself, since if it were, there should also have been forgetting instead of reminiscence in Exp. I. Instead, it looks like a forgetting of the new reward magnitude, and a regression to the earlier low level of incentive motivation (Spence, 1956). It would be parsimonious if the reminiscence in the earlier experiments could be explained in the same way. This explanation would require that *S*s in the previous experiments have been regressing between days to a level of incentive motivation appropriate to some earlier larger reward. Such an apparent regression has been noted by Collier, Knarr, and Marx (1961) in rats shifted from a higher to a lower concentration of sucrose. Two possibilities for such a reward suggest themselves: the four pellets presented in prehandling and the 9 gm. daily ration. If spontaneous regression to these larger rewards did occur, the within-days decrement would then represent gradual relearning of the current reward.

One test of the above hypothesis was to provide rats with different reward magnitudes in the first stage of training and then look for reminiscence or forgetting in the second stage. This was done in Exp. III. A variable of trial distribution was also introduced, with the prediction that between-days effects would be reduced by permitting some of the same effects to occur between the spaced trials on the same day.

Method

Subjects and apparatus.—The *S*s were 78 female albino rats, 74 days old at the beginning of training, divided randomly into six groups of 13. The runway was the same as in Exp. II. The *S*s were carried to the experimental room and detained between trials in special carrying cages, 11 × $3\frac{1}{4}$ in. and $5\frac{1}{2}$ in. high, with wooden sides and hardware cloth floors and lids.

Design.—A 2 × 3 factorial design combined two degrees of trial distribution (10–15 sec. or a minimum of 10 min.) with three reward magnitudes in Stage 1 (four pellets per trial, one pellet, or zero). The distribution variable continued through all stages of the experiment, but the magnitude variable applied only to Stage 1. In Stage 2, all *S*s received one pellet, and in Stage 3 all received extinction. Stage 1 consisted of 30 trials, and Stage 2 of 25 trials, and Stage 3 of 20 trials. In all three stages there were 5 trials a day.

Procedure.—In Stage 1, *S*s receiving four pellets reward on each trial received a daily food ration of 8 gm. instead of 9. Otherwise, except for the changes noted above, the procedure was the same as in Exp. II.

Results

Stage 1.—No learning was apparent in the no-reward groups in Stage 1, and there was no effect of distribution on speeds in the zero-reward condition. For the groups receiving some reward, a 2 × 2 analysis of variance was performed on the last day's speeds. Faster running by spaced *S*s than by massed on the last day was significant at the .01 level, with means of 3.00 and 2.12 ft/sec, F (1, 48) = 11.73, but a trend toward faster running for larger reward fell short of significance, means of 2.79 and 2.32 ft/sec, F (1, 48) = 3.27. The F for interaction was less than one.

Reminiscence was less marked than in Exp. I. Amounts and significance levels for the six groups are shown in Table 1. Combinations of groups confirm previous findings in that the two massed-rewarded groups combined are significantly positive, t (72) = 2.32, $p < .05$, but the two zero-pellet groups combined do not

TABLE 1

Reminiscence Scores for the Six Groups in Exp. III

Group	Stage 1			Stage 2			Stage 3		
	M	*SD*	t^a	*M*	*SD*	t^a	*M*	*SD*	t^a
0 Pellet Massed	.10	.34	.80	−.59	1.15	2.41	.70	.82	3.40
1 Pellet Massed	.31	.44	2.40	.52	.72	2.14	1.25	.61	6.07
4 Pellet Massed	.12	.48	.92	.38	.76	1.56	.87	.49	4.20
0 Pellet Spaced	.21	.57	1.59	−.48	.97	1.96	.54	.72	2.64
1 Pellet Spaced	−.20	.45	1.55	.04	.39	.19	1.20	.88	5.82
4 Pellet Spaced	.09	.49	.72	−.03	1.07	.13	.51	.70	2.46

Note.—Scores for Stage 1 are averaged over 5 days, for Stage 2 over 4 days, and for Stage 3 over 3 days.
[a] Comparison with zero (df = 72). A t of 2.00 is required for significance at the .05 level.

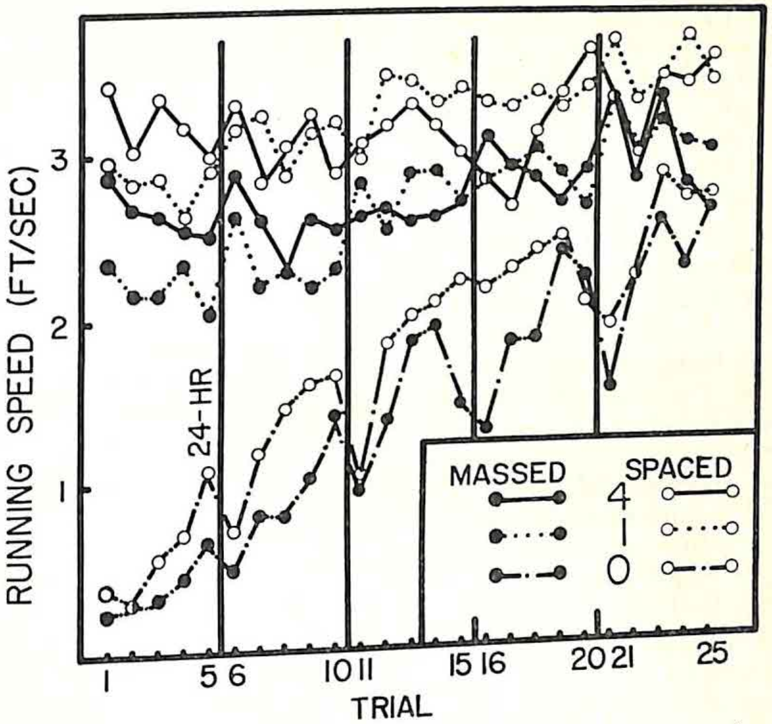


FIG. 3. Running speeds in Stage 2 of Exp. III ($N = 13$ per group).

differ significantly from zero, t (72) $= 1.70$ A repeated-measures analysis on the reminiscence scores for the six groups over the 4 scorable days revealed only one significant source of variance, the interaction of magnitude with spacing, F (2, 72) $= 3.13$, $p < .05$. However, a 2×2 analysis of variance on the four rewarded groups showed significantly less reminiscence in the spaced groups than in the massed, F (1, 48) $= 4.31$, $p < .05$, with no effect of magnitude or interaction. The two spaced rewarded groups combined did not differ significantly from zero, t (72) $= .59$.

Stage 2.—The results in Stage 2 (when all *S*s received one pellet) are shown in Fig. 3. The four previously rewarded groups appear to be approaching asymptotic speed, while the two previously nonrewarded ones show learning curves similar in rate to those of the one-pellet groups in Stage 1, indicating that there was no appreciable latent learning.

Reminiscence scores are summarized in Table 1. For the two previously rewarded massed groups, reminiscence is both significant, t (72) $= 2.62$, $p < .01$, and nonsignificantly greater than for the same groups in Stage 1, indicating that the greatly reduced slope of the learning curve does not result in any reduction in reminiscence. Both previously nonrewarded groups, in contrast, show forgetting, significant at the .01 level for the combined groups, t (72) $= 3.09$. In a repeated-measures analysis of the six groups, only the effect of prior magnitude was significantly related to the magnitude of reminiscence, F (2, 72) $= 6.60$, $p < .02$. However, when only the previously rewarded groups were analyzed, the F for magnitude was less than 1.00, while reminiscence was greater for the massed than for the spaced conditions, F (1, 48) = 4.30, $p < .05$, and the interaction remained nonsignificant.

Late in Stage 2 the two previously nonrewarded groups, like those in Exp. II, showed a suggestion of the slow first trial, fast second trial, and subsequent decline reported by Logan (1960). The effect is small, but might have become larger if rewarded training had continued.

Stage 3.—In extinction, significant reminiscence (spontaneous recovery) was present in all groups, as shown in Table 1. A repeated-measures analysis of these reminiscence scores showed significant effects of Stage 1 reward magnitude, F (2, 72) = 5.59, $p < .01$, of days, F (2, 144) = 5.35, $p < .01$, and of the interaction of days with distribution, F (2, 144) = 9.08, $p < .01$. The main effect of magnitude reflects primarily the greater spontaneous recovery in the one-pellet groups; the effect of days reflects a decline in spontaneous recovery over successive days, and the interaction reflects a steeper decline over days in the massed groups.

Discussion

The results of the previous experiments were replicated, though the effects tended to be smaller than before. In Stage 1, reminiscence was found under massed conditions, which were those used before. In Stage 2, there was forgetting between days in those groups previously nonrewarded and reminiscence in the massed groups previously rewarded.

The main hypothesis of the experiment was not supported. There was no evidence that the magnitude of reward in Stage 1, provided it was greater than zero, had any effect on reminiscence in Stage 2. The fact that in Stage 1 there was greater reminiscence with larger reward for the spaced *S*s (though nonsignificantly so) is also inconsistent with the hypothesis.

The secondary hypothesis, that spacing of trials would reduce both reminiscence and forgetting, received some support. When the zero-pellet groups were omitted from the analysis, massing gave more reminiscence than did spacing in both Stages 1 and 2. This result is consistent both with the spontaneous regression notion and with the interpretation that a dissipation of reactive inhibition is responsible for the reminiscence.

It is possible that a two-factor interpretation will be needed to account for the various effects we have noted. Perhaps the reminiscence is due to some inhibitory process and the forgetting is due to spontaneous regression of incentive motivation. The combination of the two effects could then explain the brief appearance in previously nonrewarded groups in Exp. II and III of the pattern of slow first trial, fast second trial, and decline (Logan, 1960).

The most impressive finding in this sequence of experiments is that groups receiving exactly the same current treatment show opposite between-days effects depending on whether they were rewarded or nonrewarded in an earlier block of trials. Although the explanation for this striking difference is still unclear, it points the way toward further research on a new aspect of the effect of rewards on behavior.

REFERENCES

COLLIER, G., KNARR, F. A., & MARX, M. H. Some relations between the intensive properties of the consummatory response and reinforcement. *J. exp. Psychol.*, 1961, **62**, 484–495.

ESTES, W. K. The statistical approach to learning theory. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. *General systematic formulations, learning, and special processes.* New York: McGraw-Hill, 1959. Pp. 380–491.

GLEITMAN, H., NACHMIAS, J., & NEISSER, U. The S-R reinforcement theory of extinction. *Psychol. Rev.*, 1954, **61**, 23–33.

HILL, W. F. Activity as an autonomous drive. *J. comp. physiol. Psychol.*, 1956, **49**, 15–19.

HILL, W. F. Effects of activity deprivation on choice of an activity incentive. *J. comp. physiol. Psychol.*, 1961, **54**, 78–82.

HILL, W. F., & SPEAR, N. E. Resistance to extinction as a joint function of reward magnitude and the spacing of extinction trials. *J. exp. Psychol.*, 1962, **64**, 636–639.

HULL, C. L. *Principles of behavior.* New York: Appleton-Century-Crofts, 1943.

JENSEN, A. R. On the reformulation of inhibition in Hull's system. *Psychol. Bull.*, 1961, **58**, 274–298.

JENSEN, G. D., & COTTON, J. W. Successive acquisitions and extinctions as related to percentage of reinforcement. *J. exp. Psychol.*, 1960, **60**, 41–49.

LEWIS, D. J. Acquisition, extinction, and spontaneous recovery as a function of percentage of reinforcement and intertrial intervals. *J. exp. Psychol.*, 1956, **51**, 45–53.

LOGAN, F. A. *Incentive.* New Haven: Yale Univer. Press, 1960.

RENNER, K. E. Influence of deprivation and availability of goal-box cues on the temporal gradient of reinforcement. *J. comp. physiol. Psychol.*, 1963, **56**, 101–104.

SPEAR, N. E., & HILL, W. F. Methodological note: Excessive weight loss in rats living in activity wheels. *Psychol. Rep.*, 1962, **11**, 437–438.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

(Received May 18, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 2, 210–217

PERCEPTION OF DEVIATIONS IN REPETITIVE PATTERNS[1]

CORD B. SENGSTAKE[2]

University of Oregon

This is an investigation of the stimulus variables affecting the difficulty experienced by Os in locating deviations in repetitive patterns. 11 objective variables and 7 subjective variables were correlated with the measure of stimulus difficulty, i.e., mean log time required by Os to locate the deviations. A stepwise regression was performed which indicated that 5 objective and 2 subjective variables resulted in a multiple $R = .827$ and the addition of any of the other variables did not significantly increase the predictive ability of the equation. The 7 variables that entered the regression equation were discussed in terms of some possible reasons why they were useful in the prediction of stimulus difficulty.

In the past few years, there has been a concerted effort to differentiate and evaluate some of the stimulus variables that affect visual perception. One method for exploring the visual processes is the use of the visual-search task. In this task, *O* is seeking a known stimulus object (signal) which is located somewhere in a visual field made up of greater or lesser amounts of irrelevant material (noise). The *O* must therefore seek out the signal while ignoring, as best he can, the noise which tends to obscure the signal.

In this case, the object to be found is known prior to the search. In our daily experience, however, we often are called upon to seek out the unexpected rather than the expected. Rather than knowing what it is we are seeking, in this case we know more nearly what it is that we are *not* seeking.

While it seems clear that the two types of visual-search task are not distinct entities, but rather extremes of a continuum, it may well be that the two types are worthy of separate study. At least there does seem to be a meaningful distinction between the two types of task, one requiring prior knowledge about the perceptual object or figure, the other requiring prior knowledge about the perceptual field or background.

There is still another way to contrast these two extreme types of the visual-search task, viz., lawful vs. random positioning of the stimulus parts. One type of task, that of the signal in the noisy field, can be thought of as a lawful component in an otherwise unlawful or random field. The opposite type of task can be identified as one in which an unlawful or random component is located somewhere in an otherwise lawful field.

The task of *O* under these two extreme conditions of visual search is quite different. In the case of the signal in the noisy field, the requirements are to let the unlawful field cause as little disruption as possible as the search continues for the known signal. On the other hand, when the field is the lawful structure, and the unlawfulness is the structure that is sought, then *O* is required first to determine the nature of the lawfulness, and then to determine where in the visual display there is a portion that fails to conform to this lawfulness.

---

[1] This research is based on a dissertation submitted to the University of Oregon in partial fulfillment of the requirements for the PhD degree. The investigation was carried out during the tenure of a predoctoral fellowship from the National Institute of Mental Health, United States Public Health Service. The author wishes to express gratitude to Fred Attneave for his suggestions and constructive criticisms.

[2] Now at Department of Physiology, University of California at Los Angeles Medical School.

From an experimental point of view the reversal of the relative position of the lawful and the random portions of the visual display leads to a rather interesting possibility. Because the background is ordered, it is possible to measure the amount of information contained in the background. It is true, of course, that even a random pattern can be measured for amount of information, but the informational count will naturally depend completely upon the number of units within the visual field, the number of states possible for each unit, and the probability of each of these states being assumed by each unit. A lawful pattern, in addition, has a determinable amount of redundancy; the specific amount depends upon the rules that determine the formation of the pattern. It is this redundancy that makes any given portion of the pattern determinable, assuming the rest of the pattern is known.

The present study made use of a random signal in an otherwise lawful field. The purpose was to ascertain the nature and relative importance of some of the variables that determine the difficulty, or ease, of the visual-search task.

METHOD

Stimuli

The stimuli were constructed by a modification of a method suggested by Attneave (1959). The first step was to form an $m \times m$ matrix of squares. The state of the squares was either black or white; the choice was random with each state having a probability (p) of .5. This $m \times m$ matrix will be referred to as the *basic pattern unit*. The second step was to repeat spatially the basic pattern unit to form an $n \times n$ set of such units. The *pattern* thus formed had a total of $(mn)^2$ squares. The third, and final, step in the construction of the stimuli was to pick at random one of the basic pattern unit reiterations and reverse the state of one or more of the squares within that unit. The location of the squares changed within the unit in

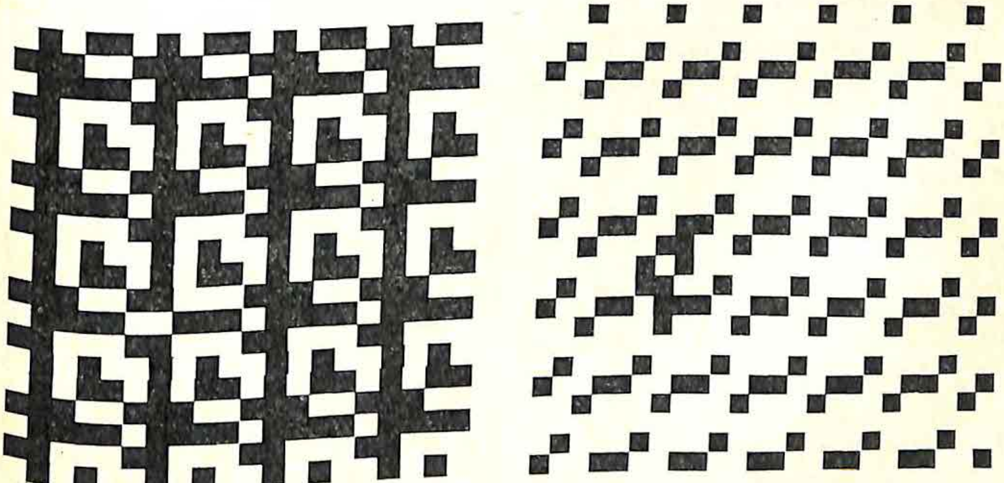


FIG. 1. Examples of the stimuli.

the formation of the deviation was also randomly determined. Any square so chosen became black if originally white, and vice versa.

The actual construction was carried out by photographing white and black, unglazed, 1-in. square tiles arranged in configurations that were determined by the above three steps. Reproductions were then made using an enlarger so that all finished stimuli were $3\frac{1}{2}$ in. square, including a small white border. Holding the overall size constant meant, of course, that the size of the individual white and black squares varied from pattern to pattern; the total number of squares in the design determined the size of the individual squares after reproduction.

Two examples of the stimuli, complete with deviations, are shown in Fig. 1.

In order to limit as little as possible the population from which the stimuli for this study were drawn, only one additional restriction was applied in determining the nature of the stimuli. The physical size of the patterns, when laid out in the tiles, soon became restrictive, as either m or n increased. The different combinations of size of each basic pattern unit ($m \times m$) and the number of reiterations ($n \times n$) that were used were as follows:

Values of m	Values of n			
	3	4	5	6
3		X	X	X
4	X	X	X	X
5	X	X	X	
6	X	X		

In each cell containing an X, five different repetitive patterns were constructed, for a total of 60 different patterns. For Exp. I, each of these patterns was used six times in the construction of the stimuli, i.e., the repetitive patterns complete with deviations. The number of deviations ranged from one to six; each pattern had the entire range in number of deviations. Thus, each pattern once contained one deviation, once two deviations, etc., up to and including six deviations. In this way the total number of stimuli for Exp. I was 360; 60 different patterns, each with the range of one through six deviations. For Exp. II, each of the same 60 patterns from Exp. I was used once, this time without deviations.

Procedure

Experiment I.—The entire set of 360 stimuli was placed in random order, and this order was maintained for all *S*s. The *S*s did not all start with Stimulus No. 1, however, but rather they were started within the series each at a randomly determined point. Each *S* was shown all of the stimuli, one at a time.

Before seeing the test stimuli, each *S* viewed a training series. This training was included because of the rapid but marked learning process preliminary work indicated to be taking place when an *O* first attempted to deal with stimuli of this type. A total of 60 training stimuli, similar to the test stimuli, was shown to *S*s. During the training period an effort was made to help *S* become as proficient as possible. In this way, the likelihood was reduced that an *S* might become significantly more proficient during the testing period and thereby cause a spurious change in the results.

The *S*s were six adults with no formal background in either psychology or art. Their specific task was to indicate the location of the area containing the deviations. There were too many stimuli for an *S* to view all of them at one sitting. Four sessions, each lasting between 1 and 2 hr., including a short break, were therefore employed for each *S*. In this way, fatigue was minimized. The mean log time required by *S*s to correctly indicate the location of the deviation area was used as the measure of stimulus difficulty. When, as occasionally happened, an *S* incorrectly identified the location of the deviations, he was told that he was wrong and should keep searching. If *S* was unable to locate the deviation area within 4 min. (this occurred a total of three times, never with the same stimulus) the trial was ended and the time of 4 min. was recorded. As is usually the case with time scores, the distributions of the raw scores from this study were skewed. The raw scores were therefore converted to logarithms for the reasons discussed by Woodworth and Schlosberg (1954).

The stimulus-difficulty measure, as determined from Exp. I, was correlated with scores on a number of pattern variables for all 360 stimuli. These variables will be explained in detail below and the ranges of values for each variable, as they were found in this study, will be given.

1. Number of Deviations. First was the number of squares whose states were reversed to form the deviation. As has been stated above, the values here ranged from 1 through 6.

2. Size of the Basic Pattern Unit. Another variable that was examined was the total number of white and black squares in the basic pattern unit, or m^2. In this study, sizes of 9, 16, 25, and 36 were included.

3. Number of Reiterations. Still another variable was the total number of reiterations of the basic pattern unit in the overall pattern. Thus, the number of reiterations was equal to n^2, or, in the present study, either 9, 16, 25, or 36.

4. Absolute Net change in Black Squares Due to the Addition of Deviations. Another variable was the absolute value of the net change in total number of black squares as a result of the addition of the deviations. Thus, if two of the squares were changed from white to black, and three other squares were changed from black to white, the absolute value of the net change would be 1. The actual values in the present study ranged from 0 to 6.

5. Ratio of Black to White Prior to Addition of Deviations. The next two variables had to do with the percentage of black and white squares in the overall pattern prior to the addition of any deviations. The first of these variables employed the ratio of black squares to white squares as a means of assessing the relative difficulty of a predominately white or predominately black pattern. In the present experiment, these values ranged from .13 to 2.60.

6. Ratio of Black to White or White to Black (whichever is the larger of the two) Prior to the Addition of Deviations. The second of these variables employed the larger of the two ratios to assess the relative difficulty of a pattern as it deviated from the state of having an equal number of white and black component squares. Here the values ranged from 1.00 to 8.00.

7. Texture. By "texture" is meant the total number of unit contours, i.e., a white square next to a black, or vice versa, both horizontally and vertically, divided by the total number of unit contours possible. The total number of unit contours possible is defined as $2mn(mn-1)$. In this study, the values found were .214–.692.

8. Dispersion of the Deviations. Also examined was the distance, measured in number of squares, from each deviation to all other deviations within each stimulus. The values ranged from 0 to 73.

9. Total Number of Squares in the Stimulus. The total number of squares was determined, of course, by the size of the basic pattern unit ($m \times m$) and by the number of reiterations ($n \times n$). Under the limitations of size that were imposed, the values ranged in this study from 144 to 625.

10. Percentage of Black Corners in the Stimulus Prior to the Addition of Deviations. Another variable was the total number of corners prior to the inclusion of any deviations divided by the total number of corners possible. The total number of corners possible is defined as $(mn-1)^2$. Given a square block of four squares, one corner occurs at the common junction of the squares, when only one square is black, or when only one square is white, or when one set of cater-cornered squares is white and the other set is black. A range of values from .380 to .842 occurred in the present study.

11. Bits of Information in the Basic Pattern Unit Prior to the Addition of the Deviations. The informational content measure used was that presented by Hillix (1960):

$$\text{Bits of information} = \log_2 \frac{n!}{r!(n-r)!},$$

where n is the total number of squares in the basic pattern unit, r is the total number of black squares, and $(n-r)$ is the total number of white squares. This function is, of course, the formula for the number of combinations transformed to a logarithmic system of base 2. The values for this factor as found in the present study ranged from 3.17 to 33.93.

Experiment II.—The entire set of 60 patterns, devoid of any deviations, was shown to a group of 20 judges. The judges, undergraduates enrolled in an introductory psychology course, were first shown all 60 patterns, one at a time, in order that they might become familiar with the type and range of the material. They were then told that they would again be shown the patterns one at a time and they were to rate each pattern along a 7-point scale for the first of the subjective variables (see below). They were again shown the 60 patterns one at a time and were required to rate them on the 7-point scale for the next variable, etc. These 20 judges were never shown the stimuli used in Exp. I, i.e., the patterns with deviations.

In all, there were seven subjective variables. These variables, or continua, were presented to the judges in the following order:

12. Simple-Complex
13. Pleasant-Unpleasant
14. Regular-Irregular
15. Compact-Spread Out
16. Jagged-Smooth
17. Easy to Remember-Hard to Remember
18. Easy to Find Error-Hard to Find Error

The judges were not specifically told to use all points on the rating scale, nor was the meaning of the names applied to the continua, e.g., simple, complex, pleasant, and unpleasant, given further clarification. The one exception to the rule of no further clarification was that it was made clear to the judges

TABLE 1

PRODUCT-MOMENT

Variable No.	2	3	4	5	6	7	8	9	10
1	.000	.000	.215**	.000	.000	.000	.893**	.000	.000
2		−.495**	−.023	.143	−.161	−.224	.240**	.465**	−.024
3			.003	−.214	.167	.232	−.127*	.465**	.144
4				−.053	.283**	−.095	.192**	−.073	.001
5					−.163	.063	.024	−.142	.025
6						−.469**	−.046	−.079	−.210
7							−.026	.005	.570**
8								.105	.018
9									.098
10									
11									
12									
13									
14									
15									
16									
17									
18									

* $p < .05$.
** $p < .01$.

what was meant by the term "error," i.e., deviation, in the last continuum.

The raw data from these judges for each of the seven continua consisted of numbers from one to seven, one number by each judge for each pattern. For each continuum, each judge's ratings for all 60 patterns were ranked. There were, of course, many ties; these ties were assigned the average of the ranks they would have had if there had been no ties. Then each pattern was assigned the sum of the ranks by the 20 judges for each continuum. These values were used in the following analysis.

RESULTS

Before substantive results are presented, the reliability of the measure of stimulus difficulty used in this study, viz., the mean log time required for *S*s to discover the correct location of the deviations, deserves some comment. The mean log times for the first three *S*s tested were correlated, for the 360 stimuli, with the mean log times for the last three *S*s tested, and the resulting correlation coefficient was .832. The Spearman-Brown prophecy formula predicted from this that for the total *N* of six *S*s the coefficient of reliability should be .908.

Table 1 shows all correlations between the criterion and the 18 variables defined in the preceding section. In determining levels of significance for these correlations a problem arose in the evaluation of the number of degrees of freedom associated with the individual coefficients. Although there was a total of 360 stimuli in Exp. I, it must be remembered that these stimuli resulted from deviations being added to only 60 different patterns. In regards to the determination of the degrees of freedom, it was possible to separate the variables into two groups. One type of variable was that which was measured after the inclusion of the deviations, viz., 1, 4, 8, and the criterion. The other type of variable included all those variables from Exp. II and those variables from Exp. I that were measured prior to the addition of any deviations. When a correlation in-

CORRELATION MATRIX

11	12	13	14	15	16	17	18	Criterion
.000	.000	.000	.000	.000	.000	.000	.000	−.437**
.982**	.638**	.367**	.118	.076	−.611**	.074	.170	.627**
−.498**	−.042	−.053	−.045	−.121	−.045	−.061	−.102	−.304**
−.061	−.072	.026	.003	−.034	.028	−.052	−.072	−.289**
.094	.159	.100	.094	.016	.021	.202	.192	.234**
−.252*	−.321*	−.205	−.170	.053	.199	−.096	−.074	−.243**
−.189	.369**	.242	.248	−.080	−.222	.149	.059	−.029
.238**	.179**	.108	.038	.021	−.167**	.022	.047	−.237**
.475**	.603**	.343**	.051	−.014	−.666**	.028	.103	.321**
−.003	.443**	.399**	.276*	.132	−.482**	.307*	.253*	.050
	.662**	.378**	.130	.063	−.623**	.072	.165	.637**
		.611**	.279*	.068	−.880**	.240	.307*	.558**
			.162	−.122	−.545**	.015	.051	.305**
				−.093	−.299*	.584**	.001	.139*
					−.295*	.721**	.947**	.115*
						−.405**	−.462**	−.481**
							.784**	.170**
								.240**

volved two variables of the first type, i.e., measures taken after the inclusion of the deviations, then $df = 360 - 1 = 359$. When correlations involving two variables of the second type were performed, then $df = 60 - 1 = 59$. When the correlation involved one variable of the first type and one of the second type, $df = 360 - 60 - 1 = 299$.

The two variables that correlated highest with the criterion were 2 (Size of the Basic Pattern Unit) and 11 (Bits of Information in the Basic Pattern Unit). They were highly intercorrelated ($r = .982$) which indicated that there was little variance not common to both. The variable with the third highest correlation involving the criterion was one of the subjective variables, 12 (Simplicity). The remaining 15 variables correlated with the criterion with coefficients that ranged, in absolute value, from .481 to .029. Most of the variables were not independent. In fact, there was real doubt in the case of some of the variables whether they actually offered any information not already given by some of the other variables. For example, in addition to 2 and 11, Variables 1 and 8 ($r = .893$), 12 and 16 ($r = -.880$), and 15 and 18 ($r = .947$) all show highly significant intercorrelations.

Table 2 shows, in summary, the results of a stepwise regression performed on the data. The seven variables have been listed in the order of entry into the regression equation. The degrees of freedom associated with the F ratios have been conservatively assumed to be (1, 59) for the first variable, (1, 58) for the second, etc. However, if a less conservative estimate of the degrees of freedom is used, e.g., (1, 359), no changes in the associated levels of confidence occur. The regression-equation coefficients are those calculated after all seven variables have been admitted to the equation. The values of R and R^2 are given after each step of the multiple regression. None

TABLE 2

STEP-WISE MULTIPLE-REGRESSION SUMMARY DATA

Variable	*F*[a]	Coefficient	*R*	*R*²
11 Bits of Information in the Basic Pattern	244.65**	.01756	.637	.406
1 No. of Deviations	169.51**	−.09298	.773	.597
12 Simplicity	31.81**	.00037	.794	.630
4 Absolute Net Change in Black Due to Deviations	23.98**	−.04935	.809	.654
5 Ratio of Black to White	23.36**	.11405	.822	.675
15 Compactness	4.50*	.00004	.824	.679
10 Total No. Black Corners	4.65*	−.26739	.827	.683
Constant		1.36886		

Standard error of estimate = 0.223

[a] *F* ratio from analysis of variance for reduction in sum of squares due to the variable entering the regression equation.

* $p < .05$.

** $p < .01$.

of the other 11 variables could significantly reduce the remaining variance associated with the criterion.

The multiple regression was used as a substitute for the experimental manipulation in order to deal with the problem of nonindependence of the variables. It would appear, then, from Table 2, that only Variables 11, 1, 12,4, 5, 15, and 10, in that order, significantly contributed in the multiple correlation. Thus, when Variables 1, 4, 5, 10, 11, 12, and 15 were held constant, the remaining variables added little if anything to the ability to explain the variance noted in the criterion.

It should be made clear that there is a difference between not accounting for any additional criterion variance and not accounting for any criterion variance at all. Consider Variables 2 and 11. Both of these variables correlated highly with the criterion; even higher was the intercorrelation between the two. The multiple-regression process picked the variable that had the highest predictive ability, as indicated by the correlation coefficients, and entered that variable into the multiple-regression equation. All remaining variables were then considered as if the variable with the highest predictive ability was held constant. That is to say, all remaining variables were considered minus any overlap that might have existed with the best predictor, as indicated by the intercorrelations between the best predictor and the remaining variables. The second best predictor with the first variable held constant was chosen in the same manner as the first, and so on.

It can be seen then that Variable 11 was the best single predictor of stimulus difficulty. Of the remaining variance associated with stimulus difficulty, Variable 1 was the best predictor. Of the variance remaining after the predictive powers of 11 and 1 were subtracted, 12 was the best remaining predictor, etc. The exclusion of Variable 2, then, was a result of the small amount of predictive power left to 2 after the removal of the predictive power common to both 2 and 11, and should not be taken to mean that 2 has no value as a predictor. What it does mean, of course, is that 2 has little further value as a predictor if 11 has already been employed as a predictor.

DISCUSSION

Variable 11 (Bits of Information in the Basic Pattern Unit Prior to the Addition of Deviations) was clearly the most important in terms of understanding what makes a task such as the one under consideration either more or less difficult. As was stated previously, a primary task associated with seeking an unknown stimulus object in a lawful visual field is the discovery of the pattern which constitutes the field. The more informational the visual field, the more difficult this task, and therefore the longer time required to find the area that does not conform.

The other variable that had relatively high predictive ability was 1 (Number of Deviations). That this variable should be rather highly correlated with the criterion is not too difficult to understand. As the number of deviations increases, the basic pattern unit containing the deviations becomes more and more different from the other basic pattern units in the stimulus. This increased difference tends to make the location of the basic pattern unit containing the deviations more obvious to *O*.

Variable 4 (Absolute Net Change in Black Due to the Addition of Deviations) and 5 (Ratio of Black to White Prior to the Addition of Deviations) both contributed to a lesser degree to the multiple-correlation coefficient. One way *O*s apparently found some of the deviations was by the rather obvious change in density, i.e., the ratio of black to white, at the site of the deviations. The perceptual result of such a change in density is that at the site of the deviations there is either a white or a black area that seems to stand out as being different from the rest of the pattern (see Fig. 1b). The higher the absolute value of the net change, the greater the likelihood that such an area would be perceptually available.

Variable 5 allows us to examine the effect of overall density of the pattern. In this case, however, we are concerned only with density in terms of the relative amount of black. The positive correlation between 5 and the criterion would indicate that as the relative amount of black increases, the difficulty of the task of finding deviations also increases. It is quite probable than an aspect of figure-ground is involved here. Since *O* must determine the nature of the pattern, it seems likely that at least part of the time he must try to organize the pattern into units. The correlation between the criterion and 5 would make sense if one considers the difficulty of trying to organize and remember the units of the pattern as black units on a white background. In this way, the greater the relative amount of black, the greater the probability that the black figure thus formed is going to be more difficult for *O* to handle. That is to say, the amount of information in the black figure increases as the number of black squares making up that figure increases.

Variable 10 (Total Number of Black Corners) also contributed slightly to the reduction of the unexplained criterion variance. The importance of the number of corners probably related back to the fact that an *O* tends to organize the black squares into perceptual units. The probability that the perceptual units will be complex, and thereby harder to deal with, increases as the number of corners increases. It would seem, then, that for at least some of the stimuli, *S*s must attempt to compare the shape of the perceptual units within the stimulus. If the shapes of the perceptual units must be compared then the more difficult will be the comparison as the number of corners of the perceptual unit increases.

Variables 12 (Simplicity) and 15 (Compactness) were the only two subjective variables to be included in the multiple-regression equation. That they were included would indicate that the objective variables failed to cover all aspects of stimuli difficulty. Thus, the judges were reacting to aspects of the stimuli that were not measured objectively.

REFERENCES

ATTNEAVE, F. Stochastic composition processes. *J. Aesthet. art Crit.*, 1959, **17**, 503–510.

HILLIX, W. A. Visual pattern identification as a function of fill and distortion. *J. exp. Psychol.*, 1960, **59**, 192–197.

WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology.* (Rev. ed.) New York: Holt, 1954.

(Received May 4, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 2, 218–223

VISUAL FIELD AND THE LETTER SPAN

HERBERT F. CROVITZ

Veterans Administration Hospital, Durham, North Carolina

AND H. RICHARD SCHIFFMAN

University of North Carolina

3 experiments tested the relations between monocular or binocular viewing, visual-field placement of stimuli, and interletter spacing in the distribution of errors over letter positions in the letter span. Exposure time was 100 msec. The most important variable in accuracy within the letter span was found to be the relative location of a letter within the 8-letter line. Errors were fewest for the leftmost letter in the line regardless of variations in absolute retinal locus. These data imply that experiments on the letter span might better be conceptualized as "memory" studies than as "perception" studies.

Edridge-Green (1914, 1920) hypothesized that the left eye dominates for the left visual field and the right eye dominates for the right visual field in normal binocular vision. Crovitz and Lipscomb (1963) reported that when the stimulus is a line of numerals and exposure time is 100 msec., accuracy is a joint function of visual-field position of a numeral and whether viewing is binocular or with the left or the right eye alone, in general accord with the hypothesized temporal visual-field dominance.

The present study investigates the perception of alphabetical material in the left or the right visual field with the same three viewing conditions. There is convincing evidence that perception in the left vs. the right visual field *for alphabetical material* is related to learned patterns of eye movement associated with reading habits (Bryden, 1960; Heron, 1957; Mishkin & Forgays, 1952). Prior to this modern interest in the letter span as a context in which to study a presumed postexposure scanning process, it was well-known that positional effects occur in the letter span. For example, Wagner found a bow-shaped error function for letters within centrally fixated eight-letter words, and Woodworth assumed that the more accurate report of peripheral letters reflects a visual-masking process in which the more isolated end letters have the advantage over the more crowded central letters which would otherwise have the advantage in virtue of their foveal location (Woodworth, 1938, pp. 742–743).

Harcum and Jones (1962) reopened the question of position effects within the letter span by analyzing the locus of errors for letter positions within eight-letter words presented wholly in the left or the right visual field. They reported a bow-shaped error function over letter position for words presented in the left visual field, but a gradually increasing error function for letters the further they were to the right for words in the right visual field.

In order to assess in a more detailed way the roles of spacing of letters and visual-field location of letters within the letter span, and to determine any differential effects of viewing condition, the following set of interlocking experiments was done.

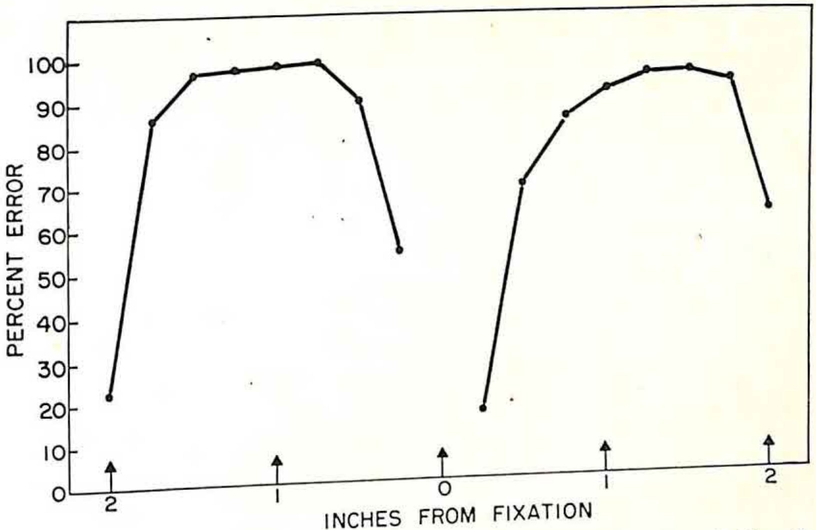


FIG. 1. Percentage of error for letters in their absolute locations in Exp. I.

EXPERIMENT I

Experiment I was designed to present a line of eight letters wholly within the left or the right visual field with the other field left blank. This is similar to the goal of the experimental design of Harcum and Jones (1962), except that they used words as stimuli. Since our goal is the assessment of letter location, per se, with minimization of the effects of cognitive factors and guessing, we used lines of eight randomly selected letters. There were three viewing conditions: left eye alone, right eye alone, and binocular viewing.

Method

Stimulus materials.—The stimuli were designed to give a line of eight letters extending into the left or the right visual field. There were 36 stimulus cards, 18 with letters to the left and 18 with letters to the right of the center. Each card contained a line of eight randomly selected letters, with the restriction that no letter appear on a given card more than once. The English capital letters were drawn with black India ink on white cards using a Zephyr lettering kit. Each letter stood $\frac{1}{4}$ in. high. The letters were spaced with $\frac{1}{4}$ in. between the centers of adjacent letters and $\frac{1}{4}$ in. between the center of the most centrally placed letter and the center of the card where the pre-exposure fixation dot appeared. When presented in a Gerbrands mirror tachistoscope the lines of letters, 26 in. from the eyes, subtended a visual angle of about 4°24′. The visual angle subtended by an individual letter was about 30′.

Procedure.—Cards were presented in a fixed random order, the two sets of cards intermixed. Exposure duration was 100 msec. There were three viewing conditions: binocular viewing, monocular viewing with the left eye, and monocular viewing with the right eye. The sequence of viewing conditions was intermixed over the set of cards, with a separate randomly selected order for each *S*.

The *S*s were instructed to maintain fixation until the letters appeared, to adjust the eye patch according to *E*'s instructions prior to each trial, and to write on an answer sheet after each trial the letters seen in the positions in which they had appeared.

Subjects.—Twenty male undergraduate volunteers served as *S*s. Prior visual screening eliminated any *S* with marked acuity differences between the eyes or marked heterophoria.

Results

Figure 1 shows the percentage of error for letters in each of the 16 visual-field loci. It can be seen that the two curves are remarkably similar; as if

TABLE 1

Analysis of Variance of Correct Report of Letters in Their Correct Relative Position by Visual Field: Exp. I

Source	*df*	*F*
Letter Position (LP)	7	44.84***
Visual Field (VF)	1	2.60
Viewing Condition (C)	2	<1
LP × VF	7	6.85***
LP × C	14	<1
VF × C	2	1.21
LP × VF × C	14	<1
Within Cells (*MS*)	912	(0.99)

*** $p < .001$.

the relative position of a letter within the line of letters determined its reportability, rather than its location in absolute visual space. Thus the results of Exp. I confirm Wagner's finding of a bow-shaped error function in the letter span. And this outweighs any difference between viewing conditions, as shown in an analysis of variance summarized in Table 1. This analysis of variance is based on a comparison of letters having the same relative position within a line, regardless of absolute location of the letter in space. *t* tests between relative letter positions in the left vs. the right visual field indicate that only Positions 2 and 3 are significantly different across visual fields. (Positions 1 through 8 shall refer to the leftmost through the rightmost letter in a line; thus Position 2 is quite far from the center of the fovea for letter lines presented in the left visual field, but quite a bit closer to the center of the fovea for letter lines presented in the right visual field.)

These bow-shaped error functions indicate that Positions 1 and 8 are most accurately reported (*t* tests show they are different from all other positions within a visual-field condition, $p < .01$).

Experiment II

The data from Exp. I indicate that the shape of the error function is similar in the two visual fields, and there is no effect of viewing condition. The question of Exp. II is whether the shape of the error function is different, and whether viewing condition matters, when fixation is in the center of the line and there is stimulation of the left and the right visual field simultaneously. To assess the

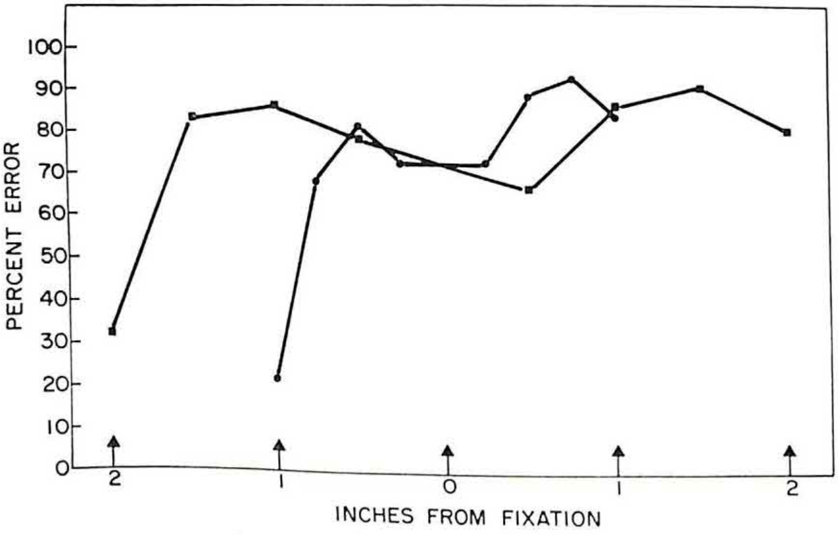


Fig. 2. Percentage of error for letters in their absolute locations in Exp. II.

hypothesis that visual masking accounts for the bow-shaped error function, spacing between adjacent letters was varied.

Method

Stimulus materials.—The stimuli were designed to give lines of letters of different spacing across the visual field. Stimuli were identical to those of Exp. I except that for 18 cards there was $\frac{1}{4}$ in. between the centers of adjacent letters, and for 18 cards there was $\frac{1}{2}$ in. between the centers of adjacent letters. In both sets of 18 cards the letters were centered on the card with a space in the center of the card equal to twice the distance between adjacent letters. For the wide-spacing cards the letter lines subtended about 8°48′ of visual angle; for the narrow-spacing cards the subtense was about 4°24′ of visual angle.

Procedure and subjects.—The procedure was the same as in Exp. I and *S*s were 29 new male undergraduate volunteers screened for visual characteristics.

Results

Figure 2 shows the percentage of error for letters in each of the visual field loci in Exp. II.

Again the two curves are remarkably similar, quite as if relative position determined report to a greater extent than did absolute locus in the visual field. And again viewing condition had no effect, as indicated in the analysis-of-variance summary shown in Table 2.

TABLE 2

ANALYSIS OF VARIANCE OF CORRECT REPORT OF LETTERS IN THEIR CORRECT RELATIVE POSITION BY SPACING CONDITION: EXP. II

Source	*df*	*F*
Letter Position (LP)	7	26.32***
Spacing Condition (S)	1	<1
Viewing Condition (C)	2	<1
LP × S	7	6.89***
LP × C	14	1.32
S × C	2	<1
LP × S × C	14	<1
Within Cells (*MS*)	1344	(1.39)

*** $p < .001$.

Within each spacing condition there was a scattering of significant differences depending on letter position, but across spacing conditions, only Relative Position 2 was significantly different. These error functions reveal an interesting feature of the interaction between relative letter position and spacing condition: the curves cross over after Position 4. A separate analysis of variance showed the interaction between spacing condition and visual field (Positions 1 through 4 pooled vs. Positions 5 through 8 pooled) was highly significant, F (1, 112) = 19.62, $p < .001$. This means that with wider spacing, the letters in the right visual field are reported relatively more accurately.

In general, however, the differences in spacing in Exp. II had little effect.

EXPERIMENT III

In this experiment the aim was to maximize the effect of a possible underlying visual-masking variable in the letter span by splitting the line of letters into two parts leaving a very wide gap between Positions 4 and 5. Here Positions 4 and 5 should be reported relatively well if accuracy in the letter span is indeed related to relative spatial isolation of the letters.

Method

Stimulus materials.—The stimuli were designed to allow a possible underlying visual-masking process to exhibit itself by a relative favoring of Positions 4 and 5. Stimuli were identical to those used in Exp. I and II except for the spacing of the letter line. On all 36 cards the line was centered on the card, but with a gap of 2 in. between the centers of letters in Positions 4 and 5. Within the first and last four letters there was $\frac{1}{4}$ in. between the centers of adjacent letters. Thus the whole letter line subtended a visual angle of about 7°42′.

Procedure and subjects.—The procedure was the same as in Exp. I and II except that the

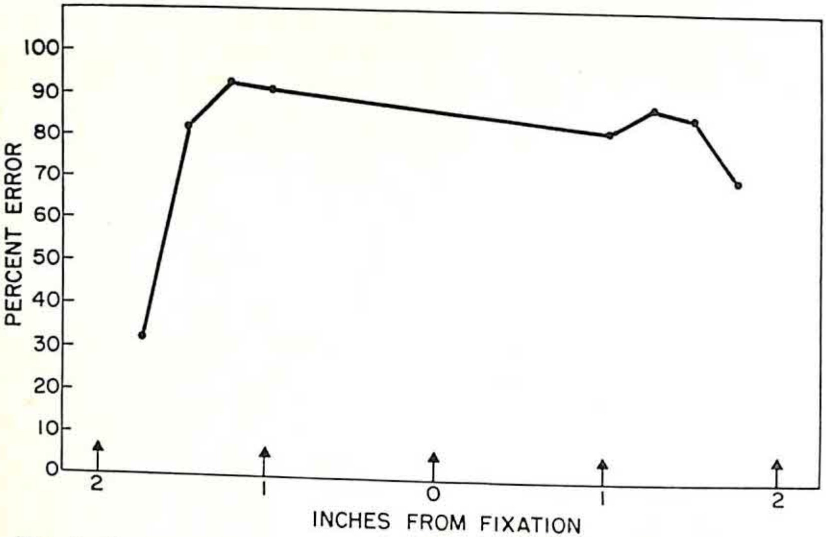


FIG. 3. Percentage of error for letters in their absolute locations in Exp. III.

same type of stimulus card was used on all 36 trials and *S*s were 20 new male undergraduate volunteers screened for visual characteristics.

Results

Figure 3 shows the percentage of error for letters in each visual field locus in this experiment. Both inspection of the curve and statistical analysis fail to find any favoring of Positions 4 and 5. And again viewing condition has no effect, as indicated in Table 3.

Positions 1 and 8 were significantly different from all other positions ($p < .01$), but Positions 4 and 5 showed no significant drop in errors as compared to Positions 2 and 3, and 6 and 7. Thus, a masking variable has failed to prove implicated in determining the shape of the error function in the letter span.

TABLE 3

ANALYSIS OF VARIANCE OF CORRECT REPORT OF LETTERS IN THEIR CORRECT RELATIVE POSITION WITH A WIDE GAP BETWEEN FIELDS: EXP. III

Source	*df*	*F*
Letter Position (LP)	7	92.67***
Viewing Condition (C)	2	<1
LP × C	14	<1
Within Cells (*MS*)	456	(3.56)

*** $p < .001$.

DISCUSSION

The results of these experiments indicate that relative position within a line of letters is the critical variable controlling accuracy within the letter span. Variations in spacing have surprisingly little effect. Further, the data do not differ depending on which of three viewing conditions is used: binocular viewing or monocular viewing with either eye alone.

Heron (1957) postulated a tachistoscopic postexposure scanning process involving habits of eye movements made in reading. In part, this process was designed to account for the facts that when a letter line appears in either the left or right visual field, accuracy is greater in the right visual field—but when letters appear in both visual fields simultaneously, accuracy is greater in the left visual field. The present data may help to clarify some details of this process. When a line of letters appears in one visual field, the error function is bow-shaped, but there is a more gradual rise in errors near the beginning of the line when it is presented in the right field. When letters appear in both fields simultaneously, the left-field letters are reported more accurately due to the remarkable accuracy of the letter in Position 1.

In Exp. I and III the error function was bow-shaped; errors were fewest in Positions 1 and 8 even when these positions were the furthest from the fovea. One possibility is that scanning occurs in a left-to-right order, from Position 1 to Position 8; that the task is fundamentally one of immediate memory in the face of rapidly decaying traces; and that both primacy and recency in the scanning order facilitate recall. This interpretation would point to similarities between the processes involved in the present experiments and those apparently involved in serial-learning tasks in which analogous error functions commonly occur (Bowman & Thurlow, 1963).

Interpretation of the results of Exp. II is more difficult. Within each spacing condition the error function is bimodal, suggesting the possibility that here the left and right halves of the letter line are scanned separately. One difference between the stimuli in this experiment and those in Exp. I and III is that in those experiments the letters were clearly segregated into left and right fields, while in Exp. II letters extended more clearly through the center of the card. Thus, here *S* might quickly come to expect letters near the fixation point; this anticipation is less likely in the other two experiments. An added complexity in the data of this experiment is the interaction between spacing condition and visual field. No hypothesis at present is available to account for this anomalous result.

Whatever the eventual explanation of details of the present findings, the data call for clarification of the processes underlying reports in "perceptual" experiments. Visual acuity, as measured by conventional tests, decreases markedly from the point of fixation outward, corresponding to a coarsening of the retinal mosaic and an increase in the ratio of convergence of cone receptors upon bipolar and ganglion cells. Accordingly, one might expect items closest to the point of fixation to be seen with greatest ease. However, these results do not show such a relation to visual anatomy. Rather, the reportability of a letter depends almost exclusively upon its *relative* position within the string of letters of which it is a member. This suggests that in the present conditions information entering through the visual system is transformed by processes closely associated with learning. Further detailed study is needed to more adequately specify the physiological and psychological processes controlling the final perceptual product in "perceptual" studies.

REFERENCES

Bowman, R. E., & Thurlow, W. R. Determinants of the effects of position in serial learning. *Amer. J. Psychol.*, 1963, **76**, 436–445.

Bryden, M. P. Tachistoscopic recognition of nonalphabetical material. *Canad. J. Psychol.*, 1960, **14**, 78–86.

Crovitz, H. F., & Lipscomb, D. B. Dominance of the temporal visual fields at a short duration of stimulation. *Amer. J. Psychol.*, 1963, **76**, 631–637.

Edridge-Green, F. W. Binocular vision. *J. Physiol.*, 1914, **48**, 38–39.

Edridge-Green, F. W. The stereoscopic appearance of certain pictures. *Nature*, 1920, **106**, 375.

Harcum, E. R., & Jones, M. L. Letter recognition within words flashed left and right of fixation. *Science*, 1962, **138**, 444–445.

Heron, W. Perception as a function of retinal locus and attention. *Amer. J. Psychol.*, 1957, **70**, 38–48.

Mishkin, M., & Forgays, D. G. Word recognition as a function of retinal locus. *J. exp. Psychol.*, 1952, **43**, 43–48.

Woodworth, R. S. *Experimental psychology.* New York: Holt, 1938.

(Received May 7, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 2, 224–230

EXPECTED VALUE AND RESPONSE UNCERTAINTY IN MULTIPLE–CHOICE DECISION BEHAVIOR[1]

DAVID M. MESSICK AND AMNON RAPOPORT

University of North Carolina

A 10-choice decision-making experiment was conducted in which the uncertainty, H, of S's response distributions was experimentally determined for each of 4 blocks of 80 trials. A measure of relative efficiency in terms of expected gain which assumes H constant, R_T, was computed for each S for each block of trials. Analysis indicated that R_T did not depend on H and that it increased slightly but significantly with blocks. The absolute values of R_T were quite close to the maximum in the later blocks, suggesting that when the uncertainty of the responses is accounted for, Ss perform the task in nearly optimal fashion. Some implications of these results for "rational" theories of decision making are discussed.

In a multiple-choice decision-making task in which an *S* is required to predict which of n stimulus events will occur on a given trial, $n > 2$, the expected value of a response is equal to the probability that the response will be correct multiplied by the payoff to *S* if the response is correct. If the probability that a response will be correct, π_i, is constant over trials for all i and if the payoff is constant, then if *S* gives a sequence of m such predictions, V, the expected value of *S*'s responses for a block of m trials, is given by

$$V = \sum_{i=1}^{n} f(i) E_i(g), \qquad [1]$$

where $f(i)$ is the number of responses with the ith alternative, and $E_i(g)$ is the expected gain of one response with alternative i.

The sequence of m responses in the n-alternative-choice task may be further characterized by a statistic, H, which is the uncertainty of the m responses. The measure H is defined as

$$H = \sum_{i=1}^{n} P(i) \log_2 P(i), \qquad [2]$$

where $P(i) = f(i)/m$, the proportion of i responses in m trials. H is a maximum where $P(i) = \frac{1}{n}$ for all i and a minimum when $P(i) = 1$ for some i.

In a previous study (Messick & & Rapoport, 1965) the V and H statistics were computed for each of 36 *S*s for five consecutive blocks of 80 trials each. In that study there were 10 alternatives ($n = 10$). Each *S* was in one of two groups, the groups differing only in terms of the payoff function determining $E_i(g)$. Within each group and block of trials the correlation between V and H was computed across *S*s. These correlations were found to be negative and to decrease monotonically with blocks. It was conjectured that the variation in H was largely responsible for the variation in V, i.e., that the value measure was increased mainly by decreasing H. The structure of this previous study, however, did not permit detailed investigation of the relationship between V and H. It is to this purpose that the present experiment is aimed.

---

[1] The authors are indebted to Lyle V. Jones for his helpful suggestions and critical reading of the manuscript. In addition the authors gratefully acknowledge the support of the Committee for the Support of Research and Advanced Study, Executive Committee, General Federation of Labor, Tel Aviv, Israel, and of the United States Air Force, Office of Scientific Research (Grant No. AF-AFOSR-85-63).

As was noted in the previous report V and H are not independent statistics. For example, the range in which V can vary is dependent on the value of H. If $H = 0$, then the range of V is maximized, while if H is a maximum the range of V is reduced to a single point. In order to account for this dependency the following simplification is introduced. Let $\Phi = \{f(1), f(2) \ldots f(n)\}$ denote the *set* of all ordered n tuples whose components are $f(1)$, $f(2) \ldots f(n)$. Thus every element of the set Φ, (ϕ), is a distinct permutation of the numbers $f(1)$, $f(2) \ldots f(n)$. It is clear that the uncertainty of every $\phi \epsilon \Phi$ is the same. It is conjectured that the uncertainty function from all sets Φ to the positive real numbers is one to one, so that it is possible to interchange uncertainty constraints and constraints imposed on the Φ's. Any particular Φ will be referred to as a response distribution.

In order to make the range of V, the value measure, independent of Φ, V is transformed into the relative measure R_T which is defined as

$$R_T = \frac{V - V_{\min}}{V_{\max} - V_{\min}}. \qquad [3]$$

In this expression $V_{\min}$ and $V_{\max}$ are extrema for a given Φ. To find $V_{\min}$ and $V_{\max}$, let Φ^* be a permutation of $f(1), f(2) \ldots f(n)$ such that $f_1 \geq f_2 \geq \ldots \geq f_i \geq \ldots \geq f_n$, where the subscript denotes position in the n tuple. Let Φ^{**} denote a permutation in which all inequalities are reversed. Let ϵ be a permutation of the $E_i(g)$ defined earlier such that $E_{i1}(g) \geq E_{i2}(g) \geq \ldots \geq E_{ij}(g) \geq \ldots \geq E_{in}(g)$, where the second subscript denotes position in the n tuple. Then $V_{\max}$ and $V_{\min}$ are given by the scalar products

$$V_{\max} = \Phi^* \cdot \epsilon, \qquad [4]$$

$$V_{\min} = \Phi^{**} \cdot \epsilon. \qquad [5]$$

The conjecture of Messick and Rapoport (1965) may now be restated to say that (*a*) R_T is independent of H, and that (*b*) R_T is roughly constant through time. If R_T is close to unity and constant and if H decreases, then V will be expected to increase. This was found in the previous study, although the assumptions regarding R_T could not be evaluated. In the present study Φ (and thus H) will be controlled experimentally to allow the assumptions regarding R_T to be evaluated in a multiple-choice, decision-making experiment.

METHOD

The *S*s were 48 undergraduate students, 30 males and 18 females, enrolled in a summer course in introductory psychology at the University of North Carolina. The *S*s were given course credit for their participation in the experiment in addition to the amount of money they won in the experiment. The *S*s, who reported individually to a dimly illuminated experimental room, were placed in three groups, A, B, and C, according to the order of their appearance (*S*s 1, 4, 7 . . . 46 in Group A, *S*s 2, 5, 8 . . . 47 in Group B, and *S*s 3, 6, 9 . . . 48 in Group C). All *S*s were seated 10 ft. from a 42 × 42 in. green projection screen and read the experimental instructions, which differed only in one word from Group C to Groups A and B.

The instructions informed *S* that his task in the experiment required him to guess how many 1's would appear in each of a set of 80 3 × 3 randomly arranged matrices composed of 0's and 1's, which would be projected onto the screen in front of him. The *S* was told that he would be given a dish containing 80 marbles of different colors and was to distribute its contents on a board with 10 grooves in front of him, putting all marbles with the same color in one groove.

The detailed procedure was explained by the following instructions.

> We shall start after you have distributed the marbles on the board. You will make your first guess of the number of *1's* to be shown on the next slide. When you make a guess you will take one marble from the column corresponding to the number you guessed, and place the marble in the empty dish. After you have guessed and removed an appropriate marble the experimenter will show you the slide and pay you 5 cents if your guess is correct. (The money you are paid will be accumulated on the desk calculator in front of you.)
>
> Note that the following processes are involved in making a correct guess:
>
> 1. The distribution of the dish's contents on the board. (You can distribute the contents as you wish, keeping one color to one groove.)
> 2. The selection of a particular marble for every slide to be shown. (You can take any marble in any order that you want.)
>
> Consider these processes so as to maximize your gain. You probably cannot guess correctly every slide but it is important for the experiment that you try to win as much money as possible. All the money you win is yours.
>
> The same 80 slides will be presented to you 6 times, being reshuffled before each presentation. The first time will be only for practice; you will only guess the number of *1's* in slides without actually using the marbles. You will not be paid on the first time. You will, however, use the marbles and be paid on the second, third, fourth, and fifth presentations of the slides. For each of these presentations, another ("the same" instead of "another" for *S*s in Group C) dish containing 80 marbles will be given to you. For the sixth presentation of the slides, you will be able to choose *freely* any of the dishes on your right (notice that the distribution of marbles is written on every dish) and to proceed as before.

The *E* answered questions and summarized the main points of the instructions. It was made clear to *S* that when marbles were gone from a given column, he would be unable to guess the number corresponding to that column. The stimuli were then mixed and placed in the carriage of a Kodak Carousel projector (Model 510). The proportion (π_i) of slides having i 1's is presented in Table 1. This is the same stimulus distribution $[g(s)]$ as was used previously by Messick and Rapoport (1965).

The experiment proceeded according to the instructions. The first presentation of the stimuli (Block I) was intended to provide *S*s of the three groups with partial information about $g(s)$, and to acquaint them with the procedure. The three groups differed, however, as to the dishes of marbles given to them by *E* at the beginning of each block, starting from Block II. The *S*s of Group A received Dishes No. 1, 2, 3, 4, in this order in the beginning of Blocks II, III, IV, V, respectively. The *S*s of Group B received the same dishes in the same blocks but in the reversed order. The *S*s of Group C received Dish No. 5 for Blocks II–V. The distributions of the marbles (Φ) in the five different dishes are presented in Columns 1–5 in Table 2 along with the $H(k)$, $V_{\max}(k)$, and $V_{\min}(k)$ for each.

Each dish fixed the response uncertainty for *S* for a given block to some predetermined value, while the allocation of the possible responses, which determines R_T, was left to *S*. From the first five entries in the row labeled $H(k)$ in Table 2, it is seen that the groups differed in terms of $H(k)$.

For the sixth block, *S*s were allowed to select any of the 13 dishes of marbles (Φ's) listed in Table 2. They were encouraged to take as much time as they wished. After *S*s made their choices of Φ the experiment was terminated and *S*s were paid.

RESULTS

The expected gain was computed separately for every *S* for Blocks

TABLE 1

PROPORTION OF SLIDES IN EACH BLOCK OF 80 TRIALS

No. of 1's in Matrix	0	1	2	3	4	5	6	7	8	9
π_i	.2250	.1375	.1000	.0750	.0625	.0500	.0500	.0625	.0875	.1500

TABLE 2

DISTRIBUTIONS OF MARBLES IN ALL 13 DISHES, THEIR UNCERTAINTY, AND THEIR MAXIMUM AND MINIMUM EXPECTED GAIN

Type of Marble (i)	No. of Dish (k)												
	1	2	3	4	5	6	7	8	9	10	11	12	13
1	17	3	1	40	25	3	12	16	23	1	1	12	3
2	14	2	30	20	1	20	9	11	18	1	49	2	2
3	12	1	24	11	20	16	20	5	5	40	21	2	1
4	10	20	11	3	16	12	16	3	3	23	4	1	60
5	8	18	5	2	3	9	14	24	1	9	2	60	9
6	7	14	4	2	4	8	4	18	28	4	1	1	5
7	6	12	3	1	3	7	3	2	1	1	1	1	0
8	6	10	2	1	2	5	2	1	1	1	1	1	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0
Σ	80	80	80	80	80	80	80	80	80	80	80	80	80
Approximate $H(k)$	2.9	2.6	2.3	2.0	2.45	2.8	2.7	2.5	2.2	1.9	1.6	1.3	1.3
$V_{\max}(k)$	52.75	58.06	65.63	71.31	62.44	55.87	57.13	61.25	65.94	71.87	76.06	80.31	80.44
$V_{\min}(k)$	27.81	24.38	23.00	22.19	23.56	26.19	25.12	23.81	22.69	21.82	21.38	21.25	20.94

II–VI. The expected gain for Block T with Dish k is given by

$$V_T(k) = 5 \sum_{i=0}^{9} \pi_i F(i,k), \qquad T = \text{II, III} \ldots \text{VI}, \quad [6]$$

where π_i was defined above, and $F(i,k)$ is the frequency of responses i given Dish k. The ratio R_T, defined in Equation 3 was computed for every S for every block. The values of R_T, which range between 0 and 1, measure how well S performed the task relative to the range $[V_{\min}(k), V_{\max}(k)]$ permitted by Dish k. The higher the value of R_T the closer is

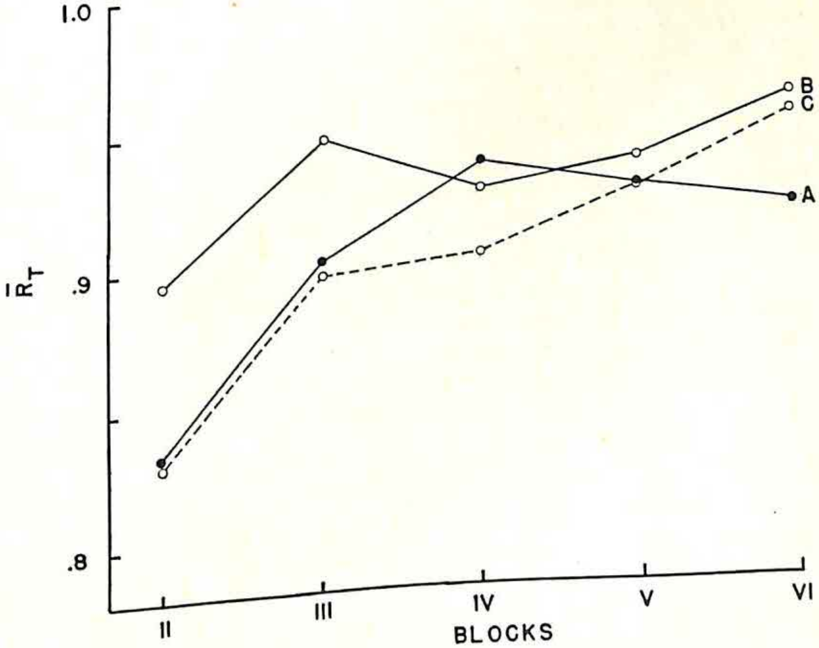


FIG. 1. Values of $\bar{R}_T$ for all groups.

TABLE 3

LOWER TRIANGULAR MATRICES OF CORRELATIONS AMONG $R_T(T = \text{II} \ldots \text{VI})$, MEANS ($\bar{R}_T$) AND *SD*s, COMPUTED SEPARATELY FOR EACH GROUP

Block	II	III	IV	V	VI	$\bar{R}_T$	*SD*
Group A							
II	1.000					.833	.1158
III	.425	1.000				.901	.0674
IV	.578*	.589*	1.000			.935	.0445
V	.447	.338	.744*	1.000		.923	.0847
VI	.417	.229	.524*	.678	1.000	.914	.0739
Group B							
II	1.000					.896	.1189
III	.408	1.000				.947	.0465
IV	.226	.715*	1.000			.925	.0800
V	−.037	.050	.320	1.000		.933	.0564
VI	.036	.312	.219	.585*	1.000	.955	.0559
Group C							
II	1.000					.830	.2033
III	.841*	1.000				.896	.1216
IV	.478	.567*	1.000			.901	.0649
V	−.015	.271	.585*	1.000		.922	.0506
VI	.073	.077	.504*	.474	1.000	.948	.0530

* $p < .05$.

S to maximizing the gain given the constraint imposed by Φ.

The values of R_T averaged over *S*s within each group, $\bar{R}_T$, are plotted in Fig. 1. These means with their standard deviations are presented in the last two columns of Table 3. It is immediately seen that the values of $\bar{R}_T$ are very high. Even for Block III, $\bar{R}_T > .89$ for all three groups. It seems that once the uncertainty of the response distribution is specified, *S*s perform the task very efficiently.

TABLE 4

ANALYSIS OF VARIANCE, TWO-FACTOR DESIGN WITH REPEATED MEASUREMENTS

Source	*df*	*MS*	*F*
Between *S*s	47		
Groups (G)	2	.0248	1.253
*S*s Within Groups (*S*s)	45	.0198	
Within *S*s	144		
Blocks (B)	3	.0559	9.164**
G × B	6	.0053	.868
B × *S*s	135	.0061	

** $p < .01$.

The summary of a repeated-measures analysis of variance of R_T for Blocks II–V is presented in Table 4.[2]

---

[2] Two points should be noted regarding this analysis. First, inspection of the correlation matrices in Table 3 suggests that the assumption of homogeneous variance-covariance structure may not be met. As a result of this possibility, the more realistic conservative tests suggested by Winer (1962) were performed. The results of these conservative tests agree with the results in Table 4. The second point concerns the logic of the test of the conjecture regarding R_T. The conjecture is a hypothesis of "no differences" and failure to reject the "no-differences" hypothesis is interpreted as support for the conjecture. When an argument is cast in these terms, as Grant (1962) has pointed out, poorly controlled, insensitive experiments are more likely to provide support for the hypothesis than are larger, well-designed studies. However, when theoretical considerations lead to an expectation of "no differences," as is the case with the conjecture discussed here, alternative procedures are not readily available. In any case confidence may be placed in the conclusion that R_T is independent of groups since $F_{.95}(2, \infty) = 3.00$, which exceeds the observed *F* of 1.253.

If, as hypothesized, R_T is independent of H, then no main effect for groups or interactions involving this factor would be expected to reach significance. This part of the conjecture is supported by the analysis. On the other hand the second part of the conjecture, that R_T is independent of time, was not confirmed. R_T appears from inspection of Fig. 1 to increase with blocks independently of the values of H encountered, reflecting the significant effect due to "Blocks" in Table 4. Note, however, that this learning effect is rather small.

For the sixth block, S was permitted to select any 1 of the 13 dishes in front of him. The median H for the combined groups was between 2.3 and 2.2. Dichotomizing the choices in terms of $H \geq 2.3$ and $H \leq 2.2$ gives the frequencies in Table 5. For this division of choices, chi-square with $df = 2$ is significant at the .05 level. It appears that Ss in Group B, in which H increased with blocks, select dishes with lower Hs than Ss in the other two groups. This finding may be interpreted as resulting from the fact that Ss in Group B in general make less money with each successive block of trials as a result of the increasing H of the given Φ. In both Groups A and C, Ss tended to earn more money as the experiment progressed. These Ss could attribute the increase in their winnings to their own improvement rather than to the H of the dish they were using, while Group B Ss, whose performance in terms of R_T increased, were compelled to find another explanation for their decreasing earnings. An obvious explanation was that S earned more with the Φ's having lower Hs than with those having the larger.

TABLE 5

NO. OF Ss ABOVE AND BELOW THE COMBINED MEDIAN

	Group A	Group B	Group C
Above	10	5	12
Below	6	11	4
	$\chi^2 = 6.60^*$		

* $p < .05$.

In addition to the investigation of variables affecting maximization of gain, another purpose of the present study was the further investigation of the stability of individual differences in maximizing gain. This variable has been largely neglected in many of the probability-matching experiments. The extent of this variable was evaluated in the present study by computing the correlations of R_T for Ss between blocks of trials. The correlations, computed separately for each group, are presented in Table 3.

The correlation matrix obtained for Group A tended to be in accord with previous findings, reported by Messick and Rapoport (1965), suggesting that response strategies become more stable as more experience is gained with the task. It is seen that for Group A the correlations between blocks increased rather regularly as a function of the block number. This was not found for Groups B or C. While Group A was run under conditions similar to those in the previous study (e.g., a decrease in H over blocks), different conditions were used for Groups B and C, which probably caused different and less stable strategies. Needless to say this interpretation is speculative and further verification of the finding is needed.

DISCUSSION

One major result of this study has been to demonstrate the feasibility of decomposing multiple-choice decision behavior into two distinct processes: one involving the determination of the relative value of the behavior as measured by R_T, and the other involving the uncertainty of the behavior as measured by H. By the very definition of R_T, these two measures are logically independent. Efficiency in multiple-choice tasks can be improved by increasing R_T (with H fixed) and/or by decreasing H (with R_T fixed and relatively high). Since the present experiment, in which H was experimentally fixed for each block of trials, is not strictly comparable to free-response experiments, the results of this study must be viewed as being suggestive of what occurs in the free-response situation.

Given this two-process conceptualization of multiple-choice decision behavior, an important finding is that R_T, the relative value measure is very high. Only 4 of the 15 means in Table 3 are below .90. If this finding can be established in free-choice experiments it will be of considerable importance to the theory of decision making, for it will imply that human response strategies are nearly optimal once H (which may be construed as a measure of response noise) is accounted for

The problem of nonoptimality of human behavior is thus transformed into the problem of response noise and its determinants. Can it be reduced or eliminated, and if so, how?

The conceptualization of multiple-choice decision behavior presented here leads directly to a procedure for making point predictions for response probabilities for all alternatives. These predictions can easily be evaluated against data obtained in free-response experiments. For a block of trials, the uncertainty of the response distribution, H, is computed by Formula 2. Assuming as before that the uncertainty function is one-to-one from the Φ's into the real numbers it is possible to associate with each H a set of "equal-uncertainty" probability vectors. Two properties of this set are obvious. First, the observed vector of relative frequencies will be a member of the set, and second, each member of the set can be obtained from any other member of the set simply by permuting the components of the vectors. (The second property of Φ_H depends on the assumption that H is one-to-one, while the first does not. If this assumption is shown to be false, point predictions can still be obtained, but it is likely that it will be necessary to program a computer to do so.) On the hypothesis that *S*s maximize expected value, given H, that permutation of the observed relative frequencies which maximizes V is selected as the predicted response probabilities. This permutation can be compared with actual data as a test of the partial maximization hypothesis.

REFERENCES

GRANT, D. A. Testing the null hypothesis and the strategy and tactics of investigating theoretical models. *Psychol. Rev.*, 1962, **69**, 54–61.

MESSICK, D. M., & RAPOPORT, A. A comparison of two payoff functions on multiple-choice decision behavior. *J. exp. Psychol.*, 1965, **69**, 75–83.

WINER, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

(Received May 11, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 2, 231–232

SUPPLEMENTARY REPORTS

PRIMACY OF A FORM CRITERION IN PERCEPTUAL JUDGMENTS[1]

ALBERT S. RODWAN
Emory University

9 Ss were each presented with 100 plane figures and told to assign them to 1 of 2 categories: "X" or "Y." These judgments were correlated with judgments of highly similar plane figures which had been assigned to "square" or "rectangle" categories. The mean correlation coefficient between the 2 judgments was .971. The 2 conclusions were: (a) height-width ratio is a prepotent cue and form is a primary judgment; (b) the use of the linear discriminant function (LDF) methodology is appropriate in discovering the subjective basis of perceptual judgments.

In a previous study Rodwan (1964) showed that *S*s produced larger *F* ratios when discriminating squares from rectangles than they did when discriminating large from small two-dimension figures. The stimuli were 100 different plane figures. The experiment was a 2 × 2 factorial analysis-of-variance design with rows being form (either square or rectangle) and columns being size (either large or small). Thus *S*s were presented with a plane figure and had to assign it to one of four categories: large rectangle, large square, small rectangle, and small square. There were two measures for each of the 100 stimulus forms, the height and the width. Eight of the 10 *S*s showed a larger *F* ratio for the rows, i.e., for the form. That is, for the same stimulus set, *S*s discriminated rectangles from squares better than large from small.

This study is an attempt to determine more directly whether or not *S*s would use form rather than size as a basis for classifying stimuli into two arbitrary categories ("X" or "Y").

Method.—The stimuli were 100 different white plane figures pasted on black construction paper and randomly arranged in two plastic loose-leaf notebooks of 50 each. There were 10 different heights and 10 different widths each ranging from 4.00 to 4.45 in. with a constant increment of .05 in. between adjacent figures. There were, therefore, 10 squares, 45 figures shorter than squares, and 45 figures taller than squares. The viewing distance varied from 24 in. to 30 in. to accommodate each *S*. Twenty-five percent of the trials were sampled and the mean viewing time was 1.15 sec.

The *S*s, five males and four females, were undergraduates at Emory University who were paid $1.00 for 1 hr. of participation.

Each *S* received two 30-min. sessions on 2 consecutive days. On the first day *S* was checked by the Bausch and Lomb orthorater and found to have at least 20/20 near vision in both eyes and was instructed informally about the experiment. The major points were:

1. This is a perception experiment.
2. Your response categories are to be "X" and "Y." You are to select your own criteria for the response categories and our only stipulation is that you remain consistent.
3. You are to look at the stimulus presented, respond with "X" or "Y" orally, and record your response on your data sheet.

Fifty practice trials were run in this session. In Session 2, Day 2, there were 150 randomized stimulus presentations, of which the first 50 were practice.

Results.—For each of these nine *S*s a linear discriminant function (LDF) was computed. The LDF requires *S* to have and use two categories which *he* defines. It gives the set of weights which *S* used in assigning inputs to these two categories. (For more detail on the LDF see Rodwan & Hake, 1964.)

The derived LDF was used to give a numerical value to each form used in this experiment; furthermore, the theoretical "form" axis derived from judgments of "square" vs. "rectangle" by Rodwan (1964) was also used to give another numerical value to each form. These two values were then correlated for each *S*. The mean correlation coefficient ($\bar{r}$) is .971 with a standard deviation of .036. Thus, unity is within one sigma unit of the mean and therefore the statement that these coefficients are unity can be accepted.

---

[1] This study was supported in part by National Science Foundation Grant No. GB-1249.

Discussion.—There are two possible functions of height and width: the height-width ratio (h/w), and the height-width product (h × w). Changing height-width ratio produces a change in shape; the height-width product, or area, is proportional to size. This study has demonstrated that the height-width ratio is a "prepotent" cue in judgments of plane figures; or that "form" is a primary judgmental criterion.

The notion of a "prepotent" cue and its related judgmental criterion is handled by Bruner's (1957) concept of access ordering. Briefly, access ordering is the ordering of judgmental criteria on the basis of either importance or likelihood of occurrence. Bruner's idea is that each individual brings to the perceptual situation certain categories which are very important to him or which he expects with a high probability to be relevant.

In summary, the results of this experiment, through the use of the LDF methodology, support the notion of the primacy of a form criterion in perceptual judgments.

REFERENCES

Bruner, J. On perceptual readiness. *Psychol. Rev.*, 1957, **64**, 123–152.

Rodwan, A. S. An empirical validation of the concept of coherence. *J. exp. Psychol.*, 1964, **68**, 167–170.

Rodwan, A. S., & Hake, H. W. The discriminant-function as a model for perception. *Amer. J. Psychol.*, 1964, **77**, 380–392.

(Received April 24, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 2, 232–234

TIME ESTIMATION AND INCREASES IN BODY TEMPERATURE

C. R. BELL

Medical Research Council Unit, London School of Hygiene and Tropical Medicine

In an attempt to replicate some early experiments which had shown consistent between-Ss changes in time estimation with increases in body temperature, Ss in the present experiments were required to (a) count to 60 at an estimated rate of 1 digit/sec, (b) tap a Morse key at an estimated rate of 3 taps/sec, and (c) judge when a metronome was beating at 4 beats/sec. Body (oral) temperatures were raised by immersion of both legs below the knees in a bath of warm water. These trials were given at each task at 3 levels of body temperature. The results of the original experiments were not confirmed.

In a previous report to this journal (Bell & Provins, 1963), data were examined for a relationship between changes in time estimation and the increase in body temperature concomitant with exposure to uncomfortably hot environmental conditions. Because no clear relationship emerged, it was decided to replicate as closely as possible the original studies of Francois (1927) and Hoagland (1933).

Method.—An experimental session lasted about 90 min. during which *S* sat in the experimental room, at an air temperature of 26° C. ± 2° C, for 10–15 min. at rest then placed his legs, to below the knee, in a bath of water at 42° C. which was then gradually raised to a comfortable temperature between 42.5° C. and 43° C. Body (oral) temperature was measured from an indwelling sublingual copper-constantan thermocouple every 2 min. and an average body temperature assigned to each period of time estimation. Instruction and practice at each of the three time-estimation tasks was given to an *S* during the first 10–15 min. after entry into the room. Data from these practice runs were not used for analyses. No knowledge of results was given to *S* throughout the experiment.

Counting task. Each of 7 *S*s was asked to tap a Morse code key at an estimated rate of one tap/sec for 60 taps. The time taken to make 60 taps was recorded to the nearest 0.1 sec. A mean of three estimations was taken and this procedure was performed three times during the experimental session.

Tapping task. Seven *S*s were each required to tap a Morse code key at an estimated rate of three taps/sec. The number of taps made in a 10-sec. period was recorded (*S* tapped for periods between 10 and 20 sec.) and a mean score calculated over four such 10-sec. periods. Three mean tapping-rate scores were collected during the experimental session.

Metronome task. The *S* was required to

TABLE 1

MEAN AND *SD*s OF BODY (ORAL) TEMPERATURES AND TIME TAKEN TO COUNT AN ESTIMATED 60 SEC., NO. OF TAPS AT ESTIMATED THREE TAPS/SEC IN 10 SEC., AND BEATS/SEC FREQUENCY PERCEIVED AS FOUR BEATS/SEC

	Oral Temp. °C.
Time Taken	
41.42 ± 15.5	37.21 ± 1.34
37.18 ± 15.5	37.45 ± 0.56
35.37 ± 15.8	37.72 ± 0.22
No. of Taps	
29.87 ± 11.3	37.18 ± 0.64
30.57 ± 6.8	37.49 ± 0.41
32.34 ± 7.1	37.68 ± 0.43
Beats/Sec	
5.38 ± 0.40	37.19 ± 0.32
5.48 ± 0.30	37.38 ± 0.45
5.41 ± 0.22	37.62 ± 0.26

stop a metronome device[1] which changed its beat between two beats/sec and eight beats/sec in either direction over 15, 20, 25, 30, or 35 sec. when he judged it to be beating at exactly four beats/sec. The metronome beat was indicated by an on/off tone and a flashing neon light. The actual frequency when each of eight *S*s made his judgment was recorded to the nearest 0.01 beats/sec. An average judgment was taken over 10 runs, 5 ascending and 5 descending, and three such average estimations were recorded during the experimental session.

Results.—Mean overall increases in body temperatures were +0.51° C., +0.50° C., and +0.43° C. (all significant at $p < .001$) on the counting, tapping, and metronome tasks, respectively. Of the corresponding changes in estimations, only the increase in speed of counting from lowest to highest body temperatures attained statistical significance ($t = 2.68$, $p < .05$). Table 1 shows the mean results on the three tasks together with the associated mean oral temperatures.

A product-moment correlation between the extent of the rise in oral temperature from lowest to highest over all *S*s and the corresponding increases in speed on the tapping task attained statistical significance ($r = +0.84$, $p < .05$). On the counting and metronome tasks, insignificant correlations were obtained of $r = -0.43$ and $r = +0.34$, respectively.

Following Hoagland (1933, 1935, 1951) the results from each individual *S* were plotted on body temperature/event frequency coordinates. It will be noted on Fig. 1 through 3 that the increase in body temperature, as expressed in degrees absolute, runs from

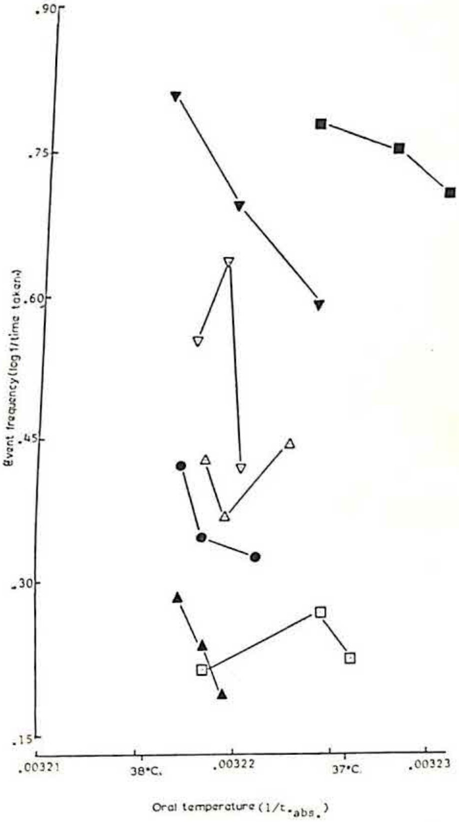


FIG. 1. Plot of speed of counting to 60 as a function of body temperature in seven *S*s (after Hoagland, 1935).

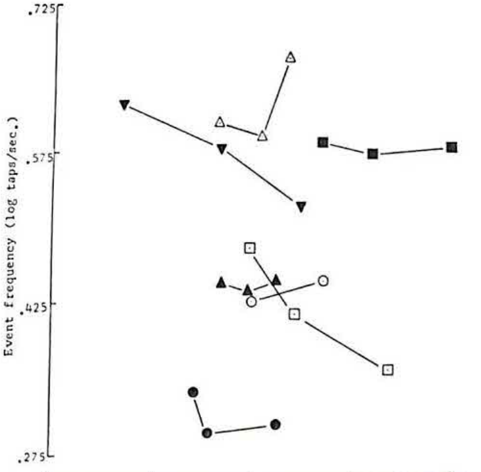


FIG. 2. Plot of estimations of tapping rate of three taps/sec as a function of body temperature in seven *S*s (after Hoagland, 1935).

---

[1] Details of this device may be had from E. S. Reeves of this Medical Research Council Unit, to whom the author wishes to acknowledge his gratitude.

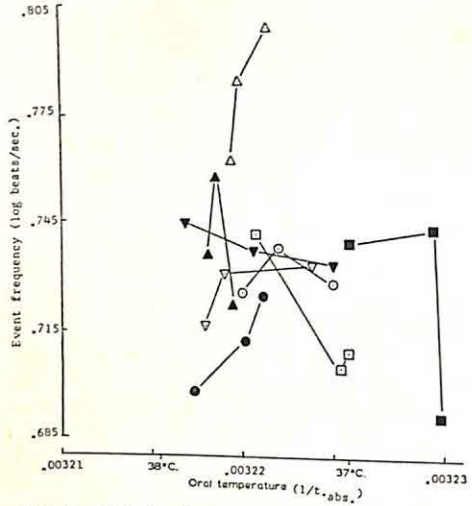


FIG. 3. Plot of judgments of metronome beat of four beats/sec as a function of body temperature in eight *S*s.

right to left and that the predicted increases in event frequency run from bottom to top. These figures confirm the considerable inter-*S* variation in average counting and tapping speeds which Hoagland found. They do not, however, confirm his finding that there is a uniformity of slope between *S*s, whether all three temperature readings are taken, as in the figures, or whether the intermediate temperature readings are ignored.

Discussion.—That there may be some influence of increases in body temperature on some aspects of time estimation is suggested by the significant increase in counting speed from lowest to highest body temperatures, and the significant positive correlation between increases in tapping speed and increases in body temperatures. It may be that the rhythmic activity aspect of time-estimation methods has a greater susceptibility to changes in body temperature than the perceptual or intellectual aspects.

It is probably never possible to replicate exactly experimental investigations conducted 30 yr. earlier by other research workers, so that differences in results may often be justifiably attributed to differences in the control of those variables which influence the data. However, until these variables have been adequately defined and controlled, the discrepancies between the conclusions of the original and present studies must stand to cast doubt on the universality of Hoagland's conclusions and on the utility of the "chemical clock" concept in the understanding of human time estimation.

REFERENCES

BELL, C. R., & PROVINS, K. A. Relation between physiological responses to environmental heat and time judgments. *J. exp. Psychol.*, 1963, **66**, 572–579.

FRANCOIS, M. Contribution à l'étude du sens du temps: La température interne comme facteur de variation de l'appréciation subjective des durées. *Annee psychol.*, 1927, **28**, 186–204.

HOAGLAND, H. The physiological control of judgments of duration: Evidence for a chemical clock. *J. gen. Psychol.*, 1933, **9**, 267–287.

HOAGLAND, H. *Pacemakers in relation to aspects of behavior.* New York: Macmillan, 1935.

HOAGLAND, H. Consciousness and the chemistry of time. In H. Abramson (Ed.), *Problems of consciousness.* New York: Corlies Macy, 1951.

(Received May 5, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 2, 234–235

GRADED CONTRAST EFFECTS IN THE JUDGMENT OF LIFTED WEIGHTS[1]

VINCENT DI LOLLO[2] AND J. H. CASSEDAY

Indiana University

6 groups of 16 Ss each judged the heaviness of either a heavy series (H) or a light series (L) of weights for either 2, 6, or 10 trials and were then shifted to the opposite series for 15 additional trials. 2 nonshifted groups acted as controls. Before the shift H was judged heavier than L. After the shift H was judged as heavier by the shift-up groups than by the H controls (positive contrast effect), and L was perceived as lighter by the shift-down groups than by the L controls (negative contrast effect). The magnitude of the contrast effects was directly related to the number of preshift trials, in accordance with predictions from adaptation-level theory.

Di Lollo (1964) demonstrated the occurrence of positive and negative judgmental contrast effects following increments or decrements, respectively, in the heaviness of lifted weights. The results were interpreted in terms of adaptation-level (AL) theory (Helson, 1959) as resulting from the remote anchoring effects of the preshift

---

[1] This research was supported in part by a research grant from the Indiana University Graduate School to the first author. We are grateful to Frank Restle for his interest in this research.

[2] Now at the University of Western Australia.

weights on postshift judgment. It may be predicted from this theory that the magnitude of the contrast effects should be directly related to the strength of the remote anchoring effects. The present study is a test of this hypothesis. The strength of the remote anchor was manipulated by varying the number of preshift trials in independent groups of *S*s.

Method.—The method has been described in detail elsewhere (Di Lollo, 1964). The *S*s were 128 undergraduate students at Indiana University. They were randomly allocated to eight groups of 16 *S*s each. Four groups judged the heaviness of a series of five light weights (L) ranging from 100 gm. to 300 gm. in steps of 50 gm. Judgments were made along a 9-point scale ranging from "very very heavy" to "very very light." Three groups were then shifted to judge a series of five heavy weights (H) ranging from 400 gm. to 600 gm. in steps of 50 gm., while the fourth group continued judging the light series. The number of preshift trials varied for each shifted groups yielding the following four conditions: LL_{10}, LH_{10}, LH_6, LH_2 (where the first letter refers to the preshift series of weights, the second to the postshift series, and the number subscript to the number of preshift trials). All groups received 15 postshift trials. On each trial *S* judged all five weights in the relevant series, randomly presented. The remaining four groups received the same treatment as the first four, with the series of weights reversed (HH_{10}, HL_{10}, HL_6, HL_2). The experiment proceeded smoothly from the preshift to the postshift stage without any interruption or warning to *S*.

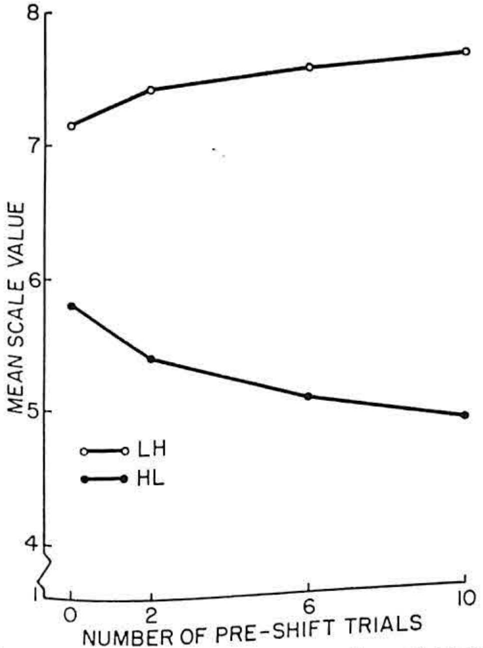


FIG. 1. Mean judgment-scale values for each group over the last eight postshift trials.

TABLE 1

ANALYSIS OF VARIANCE OF MEAN JUDGMENT VALUES OVER THE LAST EIGHT POSTSHIFT TRIALS

Source	*df*	*F*
Postshift Weight (W)	1	398.13***
Preshift Trials (T)	3	<1
W × T	3	8.21***
Error (MS)	120	(2465)

*** $p < .001$.

Results and discussion.—Each judgment was assigned a numerical value between 9 (very very heavy) and 1 (very very light), and the five responses in each trial were averaged to yield the mean judgment value for each *S* on each trial. Mean judgment values for each group averaged over the last eight trials of the postshift stage are shown in Fig. 1. The results of an analysis of variance performed on these scores are summarized in Table 1.

The occurrence of positive and negative contrast effects reported by Di Lollo (1964) is confirmed by the present results. The graded nature of the effects as a function of preshift practice, illustrated in Fig. 1 and supported by the significant interaction effect (Table 1), is in line with prediction from AL theory. A trial-by-trial plot of the mean judgment values for each group throughout the experiment was entirely similar to that reported by Di Lollo (1964). A similar plot, expressed in terms of AL values, was also like that reported by Di Lollo (1964), including a paradoxical below-zero decrement of the mean AL values following the shift in all LH Groups. The magnitude of the drop was not, however, related to the number of preshift trials.

REFERENCES

DI LOLLO, V. Contrast effects in the judgment of lifted weights. *J. exp. Psychol.*, 1964, **68**, 383–387.

HELSON, H. Adaptation level theory. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 1. New York: McGraw-Hill, 1959. Pp. 565–621.

(Received April 25, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 2, 236

Beginning with the first issue of Volume 69 and continuing through the remainder of that volume and Volume 70 (1965), the titles and authors of accepted papers will be listed here following Supplementary Reports. It is being supported on an experimental basis by the APA Project on Scientific Information Exchange in Psychology, and at the end of the year, the outcome of this trial will be evaluated and consideration given the advisability of continuing the listing.

This listing plus those published in the preceding 1965 issues are a portion of the backlog of manuscripts accepted by this journal. Such listing will allow readers to become aware of research many months in advance of journal publication.

Manuscripts Accepted for Publication in the *Journal of Experimental Psychology*

Selection Strategies in Concept Attainment as a Function of Number of Relevant Problem Attributes: Patrick R. Laughlin*: Department of Psychology, Northwestern University, Evanston, Illinois 60201.

Right-Response Preference in Probability Learning and Reversal: Marilyn E. Miller*: Department of Psychology, University of Wisconsin, Milwaukee, Wisconsin 53211.

Effect of Threshold Reduction on the Vibrato: J. K. Clarkson and J. A. Deutsch*: Department of Psychology, New York University, Washington Square, New York, New York 10003.

Mediated Association in the Paired-Associate Learning of Children Using Heterogeneous and Homogeneous Lists: David S. Palermo*: Department of Psychology, Pennsylvania State University, 117 Burrowes Building, University Park, Pennsylvania 16802.

Range of Association Level (AL) and Observing Response (OR) Effects in Postshift Concept Attainment: Roy Lachman*: Department of Psychology, State University of New York at Buffalo, Buffalo, New York 14214.

Clustering in Free Recall as a Function of Certain Methodological Variations: Charles N. Cofer,* Darryl R. Bruce, and Gerald M. Reicher: Department of Psychology, Pennsylvania State University, 117 Burrowes Building, University Park, Pennsylvania 16802.

Acquisition of Hierarchical Control over the Temporal Organization of a Skill: Richard W. Pew*: Human Performance Center, Department of Psychology, University of Michigan, Ann Arbor, Michigan 48104.

Effect of Stimulus Pattern on Temporal Acuity in the Somatosensory System: William R. Uttal* and Madelson Krissoff: Mental Health Research Institute, University of Michigan, Ann Arbor, Michigan 48104.

Temporal Integration: The Relative Value of Rewards and Punishments as a Function of Their Temporal Distance from the Response: K. Edward Renner*: Department of Psychology, University of Pennsylvania, 106 College Hall, Philadelphia, Pennsylvania 19104.

Operant GSR Conditioning and Associated Electromyogram Responses: David G. Rice*: Department of Psychiatry, Medical School, University of Wisconsin, 1300 University Avenue, Madison, Wisconsin 53706.

Effects of a Ready Signal Upon Eyelid Conditioning: Barbara B. Turner*: 2062 Chrysler Drive, N. E., Atlanta, Georgia 30329.

Relation between Stimulus Presentation Time, Serial Learning, and the Serial-Position Effect: Gloria J. Fischer*: Division of Social Sciences, University of South Florida, Tampa, Florida 33620.

Vibrotactile Thresholds for Hairy Skin: Ronald T. Verrillo*: Laboratory of Sensory Communication, Syracuse University, Syracuse 10, New York.

Absolute Judgment of Distance as a Function of Induced Muscle Tension, Exposure Time, and Feedback: N. McK. Agnew,* Sandra Pyke, and Z. W. Pylyshyn: Department of Public Health, Psychological Research Centre, 1214 College Drive, Saskatoon, Saskatchewan, Canada.

Recovery in the Absence of Recall: An Investigation of Color-Word Interference: Stanley Grand and Sydney J. Segal*: Research Center for Mental Health, New York University, Washington Square, New York, New York 10003.

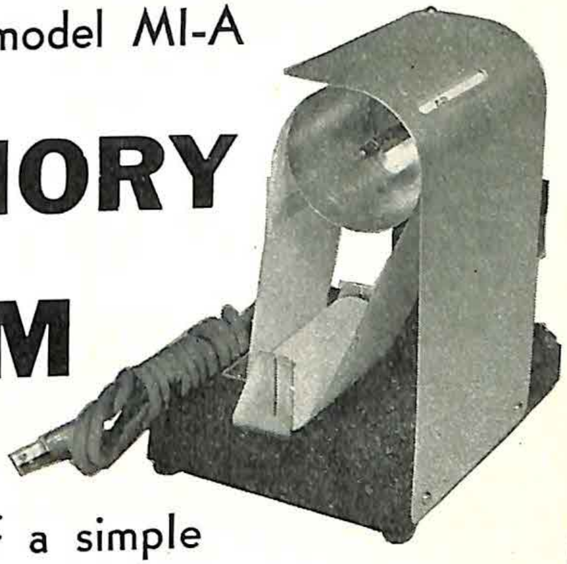
---

* Asterisk indicates author for whom address is supplied.

our model MI-A

MEMORY

DRUM



● is of a simple fool proof design, has 3 exposure "speeds", can be used with long or short word lists, is an ideal drum for use in laboratory class and sells for the low price of

$57.00 F.O.B. ARLINGTON MASS.

- FOR COMPLETE DETAILS WRITE OR PHONE

Ralph Gerbrands Company
Scientific Instruments

8 BECK RD., ARLINGTON 74, MASS. • TEL. 617-648-6415

the monthly review journal
of the
American Psychological Association

CONTEMPORARY PSYCHOLOGY

reviews

criticism

opinion

on

the latest books, films,

and other instructional media

lists of new books

letters-to-the-editor

$1.00 a copy

$10.00 a year
($10.50 foreign)

Order from:

AMERICAN PSYCHOLOGICAL ASSOCIATION
1200–17th Street, N.W.
Washington, D.C. 20036

Journal of
Experimental Psychology

VOL. 70, NO. 3 SEPTEMBER 1965

UNLEARNING UNDER CONDITIONS OF SUCCESSIVE INTERPOLATION[1]

LEO POSTMAN

University of California, Berkeley

This study investigates unlearning of 1st-list associations as a function of the number of different new responses attached to the original stimuli during a fixed period of interpolated learning. In an RI design the work groups received 16 interpolated trials which were distributed over 1, 2, or 4 different lists. Retention was measured by unpaced free recall and by S-R matching. By both measures RI increased with the number of interpolated lists. The results support the conclusion that the frequency of unreinforced evocations of original responses, and hence the amount of unlearning, vary inversely with the degree of dominance attained by the interpolated habits.

When successive lists of paired associates conform to the A-B, A-C paradigm of negative transfer, first-list associations are unlearned during the acquisition of the second list. Tests of modified free recall (MMFR), which minimize the role of response competition and list differentiation, show that the association of the first-list responses to the environmental context as well as to the specific stimulus terms is weakened during second-list learning (Barnes & Underwood, 1959; McGovern, 1964). On the assumption that unlearning has the functional characteristics of extinction (Briggs, 1954; Melton & Irwin, 1940; Underwood, 1948) the amount of unlearning should depend on the frequency with which first-list responses are elicited without reinforcement during interpolated learning (IL). Some support for this assumption was obtained in an earlier study (Postman, Keppel, & Stark, 1965) which investigated unlearning as a function of the similarity of the response classes in the successive lists. The probability of first-list responses occurring as explicit or implicit errors during the acquisition of the second list, and hence the degree of unlearning, were expected to be greater when the response class remained the same than when it changed. The differences in the amounts of retroactive inhibition

---

[1] This research was carried out under Contract Nonr 222 (90) between the Office of Naval Research and the University of California. Reproduction in whole or in part is permitted for any purpose of the United States Government. The work was performed at the Institute of Human Learning, which is supported by a grant from the National Science Foundation.

(RI) as measured by an unpaced test of free recall bore out this expectation.

The frequency with which old habits are elicited and subjected to extinction should depend not only on the similarity of the successive response classes but also on the degree of dominance achieved by the new habits. As the new response to a given stimulus gains in strength, it becomes less likely that the old response will occur as an explicit or implicit error. During a fixed period of IL, the level of dominance attained by the new habits will vary inversely with the number of different new responses attached to each of the old stimuli. Thus, the larger the number of different responses, the longer should the original response continue to be elicited as an error during IL. It follows that the degree to which first-list associations are unlearned should increase with the number of new lists practiced during a given period of IL. This hypothesis is tested in the present experiment.

The predicted results would be consistent with known facts concerning the dependence of RI on the degree and amount of IL. (*a*) Increases in RI as a function of the number of trials on a single interpolated list are rapid at first but level off thereafter (Briggs, 1957; Melton & Irwin, 1940; Thune & Underwood, 1943). The same relationship obtains in MMFR (Barnes & Underwood, 1959) so that extinction of first-list associations remains incomplete even after a very high degree of IL. According to the present interpretation, RI will increase with the number of trials in IL as long as first-list responses continue to be elicited as errors and thus are subjected to extinction. However, the rate of such unreinforced elicitations will diminish progressively as the interpolated responses become dominant. (*b*) A finding which bears directly on the present problem was reported by Underwood (1945) who compared the effects on RI of successive interpolations of different lists and of equivalent increases in the number of trials on a single interpolated list. With recall measured by the method of anticipation, the addition of new lists produced larger increases in RI than did continued practice on a single list. The present experiment is designed to determine whether such a divergence between the two conditions of interpolation reflects differences in the amount of unlearning.

METHOD

Design.—The conventional design for the measurement of RI was used. There were three work conditions and a control condition. After learning a first list of paired associates to a criterion of 10/12, the work groups were given 16 trials of IL in which new responses were attached to the old stimuli. The three work groups differed with respect to the number of different lists used in IL. In Cond. IL-1 a single list was practiced for the entire 16 trials, in Cond. IL-2 there were two successive lists given 8 trials each, and in Cond. IL-4 there were four successive lists with 4 trials on each. The control *S*s (Cond. IL-0) learned and recalled only the first list. Two methods of measuring retention were used: (*a*) MMFR, and (*b*) S-R Matching, in which *S* is supplied with a list of the correct responses and is required to match them with the appropriate stimuli. For purposes of assessing the relationship between the two methods, one half of the *S*s in each condition was tested by MMFR first and by S-R Matching second (Successive Test groups); the other half was tested by S-R Matching only (Matching-Only groups). With four conditions of training and two procedures of testing, the total design comprised eight treatment combinations.

Materials.—There were five lists of paired associates (the number required for Cond. IL-4). Each list consisted of 12 pairs, with CVC trigrams as stimuli and two-syllable adjectives as responses. The same stimulus terms were used in all lists. The trigrams had association values from 45 to 55% (Archer, 1960). Intralist similarity of the stimulus terms was kept at a minimum. There were no repetitions of consonants in the first positions or the third positions. Sixteen consonants were used once, and four consonants were used twice, once in the first and once in the third position. Five vowels and the letter *y* were used twice each in the middle positions.

In the selection of adjectives from available lists (Haagen, 1949; Melton & Safier in Hilgard, 1951; Thorndike & Lorge, 1944), every effort was made to minimize both meaningful and formal similarity within lists and between lists. Initial letters were not duplicated in any of the lists. Responses attached to the same stimulus in successive lists not only had no apparent meaningful relationship but also had different initial letters and different final syllables.

The combinations and sequences of lists were balanced so that each of the five lists (*a*) was used equally often as the first list, (*b*) appeared equally often in each of the positions of interpolation used in a given condition, and (*c*) preceded and followed every other list equally often.

Procedure.—The lists were learned by the anticipation method at a 2:2-sec. rate, with an intertrial interval of 4 sec. Four different orders of presentation were used to minimize serial learning. Each of these orders was used as a starting order equally often.

The first list was learned to a criterion of 10/12. After attainment of this criterion, all work groups were given the same instructions for IL. The *S*s were informed that the stimuli would remain the same but that the responses would be new and might change again. If such a further change were to occur, it would be announced by *E* (during the intertrial interval). The *S*s were requested to guess what the new words might be on the first trial of each new list. This procedure was introduced in order to equalize the opportunities for elicitation of first-list responses under all conditions of interpolation. In Cond. IL-2 and IL-4 it was necessary to shift the window of the memory drum from the right-hand to the left-hand side between the eighth and ninth trials of IL. The same procedure was used in Cond. IL-1 in order to insure uniformity of procedure. For this purpose an intertrial interval of 30 sec. was introduced between Trials 8 and 9. Practice on the interpolated material began 4 min. after the end of List 1 learning. A series of arithmetic problems was solved by the control groups during the period occupied by IL for the work groups. The total retention interval was 18.5 min. for all groups.

The MMFR test was administered in written form. The stimulus terms were typed in a column; next to each stimulus there were as many blanks as there had been different responses paired with it, i.e., one blank for Cond. IL-0, two blanks for Cond. IL-1, etc. The order in which the stimulus terms were arranged was the same as would have been used on the next anticipation trial on List 1 if learning had continued. Thus, there were four different forms of each MMFR test. The *S*s were instructed to write the responses paired with each stimulus in the appropriate blanks in whatever order occurred to them. At the bottom of the recall sheet blanks were provided for words which *S*s could recall but were unable to assign to a specific stimulus. A period of 5 min. was allowed for MMFR, which was fully sufficient for completion of the test under all conditions.

The S-R Matching test was also written. The stimulus terms and the lists of responses were typed in parallel columns. For the work conditions the successive lists of responses were arranged in order of acquisition. The stimulus terms were presented alphabetically as were the responses in each list; the latter were also numbered 1–12. For each list there was a column of 12 lines in which *S* filled in the numbers of the responses matching the stimulus items. The test was self-paced.

For the Successive Test groups the matching procedure followed immediately after the completion of MMFR. The interval between the end of first-list learning and the matching procedure was exactly the same for the Matching-Only groups as for the Successive Test groups. During the 5-min. interval used for the MMFR test in the successive procedure, the Matching-Only groups solved a series of mathematical problems. This arrangement made it possible to evaluate the effect of a preceding MMFR test on matching performance.

Subjects.—A total of 160 *S*s, 20 in each of the eight treatment combinations, served in the experiment. The *S*s were assigned to conditions in blocks of eight, with one *S* from each treatment combination per block. The running order within blocks was determined by means of a table of random numbers. Assignments to combinations and sequences of lists and starting orders were also randomized, with the restriction that these assignments be balanced within each treatment combination. There were no restrictions on selection of *S*s, most of whom had participated in other studies of verbal learn-

TABLE 1

MEAN NUMBERS OF CORRECT ANTICIPATIONS IN IL

List	Cond.		
	IL-1	IL-2	IL-4
2	130.4	42.8	11.2
3	—	43.4	9.8
4	—	—	10.6
5	—	—	10.3
Total	130.4	86.2	41.9

ing. No *S*s were lost because of failure to learn.

RESULTS

First-list learning.—The mean number of trials to criterion on List 1 for the combined groups was 10.0 ($SD = 5.2$). The means of the eight individual groups ranged from 7.8 to 11.8, but these variations are not significant. After log transformation, which was used because of the skewness of the distributions, F (7, 152) = 1.30.

Interpolated learning.—Table 1 presents the mean numbers of correct responses on each of the interpolated lists learned by a given work group as well as the totals for the entire period of IL. The results for the Successive Test groups and the Matching-Only groups were closely similar in each case and have been combined. The total numbers of correct responses show large and significant decreases as a function of the number of interpolated lists, F (2, 117) = 136.36, $p < .001$. These decreases reflect the differences between continued practice of old associations and acquisition of new associations under conditions of negative transfer. Moreover, the opportunities for correct responding declined with the number of trials—1, 2, and 4, respectively—required for the presentation of new sets of responses. However, the differences in the total numbers of correct anticipations substantially exceed the variations in opportunities.

In Cond. IL-2 and IL-4 there are only minor fluctuations in the numbers of correct responses on successive interpolated lists. Improvements due to learning to learn are negatively accelerated (e.g., Thune, 1951) and in the present case are also counteracted by the cumulative effects of negative transfer. The net result is a stable level of performance over the successive phases of IL.

Interlist intrusions.—The frequencies of List 1 responses occurring as intrusions during IL were as follows, with the numbers of *S*s contributing them given in parentheses: 53 (15) for IL-1, 43 (17) for IL-2, and 55 (17) for IL-4. The percentages of the total frequencies which were given to the appropriate stimulus terms were 57, 58, and 69, respectively. The low absolute frequencies of interlist intrusions are characteristic of the A-B, A-C paradigm. Less than half the *S*s in each condition of interpolation gave any intrusions at all; the mean number of intrusions during the entire period of IL is of the order of 1 per *S*. These results indicate that the level of list differentiation was high and that *S*s were successful in recognizing and rejecting incorrect responses.

There are no appreciable differ-

TABLE 2

PERCENTAGES OF INTERLIST INTRUSIONS FROM DIFERENT LISTS IN COND. IL-4

Source	Place of Occurrence			
	List 2	List 3	List 4	List 5
List 1	100	33	22	14
List 2	—	67	14	5
List 3	—	—	63	20
List 4	—	—	—	62

TABLE 3

MEAN NUMBERS OF CORRECT RESPONSES IN MMFR

Cond.	Responses									
	List 1		List 2		List 3		List 4		List 5	
	Str.	Len.	Str.	Len.	Str.	Len.	Str.	Len.	Str.	Len.
IL-0	10.6	10.8	—	—	—	—	—	—	—	—
IL-1	6.6	7.1	11.9	12.0	—	—	—	—	—	—
IL-2	5.3	5.7	5.8	6.2	9.8	10.2	—	—	—	—
IL-4	4.3	4.7	2.9	3.4	3.4	3.9	4.4	4.9	7.0	7.7

Note.—Str. = stringent scores; Len. = lenient scores.

ences in the numbers of intrusions among the work conditions. On the basis of relative associative strength alone, List 1 intrusions should increase with the number of lists. However, list differentiation increases with the disparity in strength between competing response systems (Underwood, 1949), and on that basis the probability of overt List 1 intrusions should vary inversely with the number of interpolated lists. The small variation in the actual frequencies of such errors points to a balance between the effects of associative strength and list differentiation. If this analysis is correct, it may be assumed that the frequency of implicit List 1 intrusions increased progressively with the number of lists.

Evidence for the sustained competitive strength of List 1 responses is provided by their persistence as overt intrusions throughout IL. In Cond. IL-2, 31% of all intrusions during the acquisition of List 3 were List 1 responses. The relative persistence of intrusions from List 1 and from subsequent lists can be compared directly for Cond. IL-4. Table 2 shows for the successive interpolated lists the distribution of intrusions, i.e., the percentages that responses from each of the preceding lists are of the total frequencies. In each case the highest percentage of intrusions is from the immediately preceding list. This pattern indicates, in agreement with earlier findings of Underwood (1945), that level of differentiation varies inversely with the degree of contiguity of the lists in acquisition. It is also apparent that responses from List 1 are more persistent as intrusions than responses originating from other lists. Thus, with degree of contiguity held constant, List 1 remains a more likely source of intrusions than the other lists. For example, 33% of the intrusions during the acquisition of List 3 come from List 1, but only 14% of the intrusions during List 4 learning come from List 2. When the conditions of differentiation are comparable, differences in the associative strength of competing responses are reflected in the frequency of intrusions.

Recall.—Table 3 shows the mean numbers of correct responses on the MMFR test. The stringent scores are based on responses given to the appropriate stimulus; the lenient scores are based on all responses recalled from a given list, regardless of whether or not they were given to the appropriate stimulus. The stringent scores for List 1 responses provide the basic measure of differences in recall as a function of the conditions of interpolation. These scores (*a*) are substantially lower for all work groups than for the control group, and (*b*) decline monotonically as a function of the number of interpolated lists. Analysis of variance (following a Freeman-Tukey square-root transformation to remove heterogeneity of variance) shows the differences among the conditions to be highly significant, F (3, 76) = 29.56, $MS_w = .73$, $p < .001$. Orthogonal comparisons yield reliable differences between the control condition and the combined work conditions on the one hand, and among the three work conditions on the other, F (1, 76) = 75.45 ($p < .001$) and F (2, 76) = 6.62 ($p < .01$), respectively. Thus, the amount of RI varies significantly with the number of interpolated lists. The lenient scores are slightly higher than the stringent scores but the increases are comparable for all conditions and leave the pattern of differences among conditions unchanged.

There is no apparent forgetting for the responses of the final interpolated list. The mean numbers correct on the last interpolated trial were 11.7 for IL-1, 9.0 for IL-2, and 5.8 for IL-4. These values may be compared with stringent MMFR scores for the final list of 11.9, 9.8, and 7.0, respectively. The increases between acquisition and recall may be attributed to learning on the last training trial and to the advantage of unpaced MMFR over paced anticipation. In Cond. IL-2 and IL-4 the numbers of interpolated responses recalled increase with the ordinal position of the list. While these trends probably reflect the differential opportunities for unlearning of the successive lists, they cannot be so interpreted because of the concomitant variations in the length of the retention interval. As in the case of List 1, the lenient scores are slightly higher than the stringent scores but leave the pattern of results unchanged.

Errors in recall.—When the criteria of stringent scoring are applied, two types of intralist errors can occur in MMFR: (*a*) misplaced responses, i.e., responses from within the list given to an inappropriate stimulus, and (*b*) unassigned responses, i.e., those given at the end of the test which *S* was unable to connect with a specific stimulus. The frequencies of misplaced responses were as follows, with the numbers of *S*s contributing them given in parentheses: 8(5) for IL-0, 7(5) for IL-1, 4(4) for IL-2, and 5(5) for IL-4. The corresponding values for unassigned responses were 0(0), 5(4), 6(4), and 3(3). The numbers of both types of errors are uniformly small, and there are no apparent systematic trends.

Matching.—The mean numbers of items correct on the S-R Matching test are presented in Table 4. The scores of the Matching-Only groups and the Successive Test groups are shown. While the absolute level of performance on List 1 is considerably higher than in MMFR, the matching scores again decline progressively as a function of the number of interpolated lists.[2] An overall analysis of variance (following a Freeman-Tukey square-root transformation to eliminate heterogeneity of variance) shows the differences among the conditions of interpolation to be highly significant, F (3, 152) = 15.61, $MS_w = .51$, $p < .001$. For the difference between the control and combined work conditions F (1,152) = 40.35, $p < .001$;

---

[2] It was not considered necessary to correct the matching scores for guessing since the main interest was in the differences among the conditions of interpolation. A large difference between the levels of MMFR and matching would, of course, remain after such a correction.

TABLE 4

MEAN NUMBERS CORRECT ON S-R MATCHING TEST

Cond.	Responses									
	List 1		List 2		List 3		List 4		List 5	
	MO	ST	MO	ST	MO	ST	MO	ST	MO	ST
IL-0	11.7	11.4	—	—	—	—	—	—	—	—
IL-1	9.7	9.7	11.8	11.8	—	—	—	—	—	—
IL-2	8.7	9.1	9.6	9.0	11.1	10.4	—	—	—	—
IL-4	8.2	8.9	6.3	6.5	6.3	6.7	6.9	7.4	8.4	8.2

Note.—MO = Matching-Only groups; ST = Successive Test groups.

for the variation among the work conditions $F(2, 152) = 3.25, p < .05$.

The differences between the two procedures of testing are small, but they appear to follow a systematic trend. While the Matching-Only group has a very slight advantage in Cond. IL-0, there is a progressive shift in favor of the Successive Test groups as a function of the number of interpolated lists. As a net result, the variation among the work conditions is somewhat more pronounced for the Matching-Only groups than for the Successive Test groups. However, the overall difference between the two test procedures is not significant and also does not interact reliably with the number of interpolated lists ($F < 1$ in both cases). Under both procedures the numbers of correct matches for the interpolated lists show the same trend as the MMFR scores, i.e., performance improves with recency of acquisition.

Relationship between recall and matching.—An item analysis was carried out to determine the relationship between recall and matching when these two tests were taken in succession. The conditional probabilities of a correct match, given recall in MMFR, are .995 for IL-0, .985 for IL-1, .943 for IL-2, and .942 for IL-4. The conditional probabilities of a correct match, given failure to recall in MMFR, are .643, .596, .604, and .630, respectively. Once an item has been recalled, it is almost certain under all conditions to be matched correctly as well; there is only a very slight decline in the conditional probabilities as a function of the number of interpolated lists. Similarly, there are only minor variations in the probabilities of a correct match following a failure of recall. Thus, the relationship between the two measures of retention remains nearly constant under different experimental treatments.

DISCUSSION

Retroactive inhibition in MMFR and S-R matching was found to increase with the number of interpolated lists, i.e., the retention loss varied inversely with the number of correct responses during a fixed period of IL. This relationship had been predicted on the assumption that the greater the degree of dominance achieved by interpolated responses, the less frequent would be the unreinforced elicitation of the original associations. Thus, the development of strong new habits appears to protect the old ones from continued extinction. Since these results were obtained on tests which minimize the effects of response competition and list differentiation, the variations in RI are likely to reflect differences in the amount of unlearning.

The analysis of interlist intrusions during IL gives only indirect support to the assumption that the development of strong new habits protects old responses from being elicited as errors. In spite of the fact that list differentiation typically increases with the difference in strength between competing responses, the frequency of interlist intrusions failed to decline with the number of interpolated lists. In addition, List 1 responses were more persistent as intrusions under conditions of multiple interpolation than were responses from other lists. However, the absolute frequencies of overt intrusions were extremely low, as they usually are in experiments using the A-B, A-C paradigm with familiar words as responses. It is reasonable to suppose, therefore, that the large majority of intrusions was implicit rather than explicit. The numbers of overt intrusions thus provide only very limited information about nonreinforced occurrences of List 1 responses during IL.

Performance in S-R matching is substantially higher than in MMFR, but the two methods of measuring retention yield entirely consistent trends. Correct performance in MMFR depends on both the availability and the correct placement of responses, i.e., it requires that the associations with the specific stimulus terms as well as with the environmental context remain intact. Success in S-R matching is exclusively a function of the strength of specific associations. Both classes of associations are influenced by the conditions of interpolation in parallel ways. The difference in level of performance between the two tests indicates, however, that the availability of responses declines more rapidly than the retention of specific associations. The differential sensitivity of the two tests is reflected in the relationship between recall and matching which obtained for individual items under conditions of successive testing. The responses matched correctly included not only virtually all those recalled in MMFR but also more than half of those which had not been recalled.

The rehearsal provided by the MMFR test did not result in reliably higher matching scores for the Successive Test groups than for the Matching-Only groups. The responses given in MMFR were those most resistant to extinction and were likely to be matched correctly regardless of prior rehearsal. Misplaced responses in MMFR were rare, so that the persistence of errors is not likely to have counteracted to any substantial degree positive effects of rehearsal of correct responses. An interpolated test is most likely to influence subsequent performance if it permits the rehearsal of associations of marginal strength which are liable to be forgotten without such additional practice. There does, in fact, appear to be a systematic trend toward an increasing advantage of the Successive Test groups over the Matching-Only groups as a function of the number of interpolated lists. This trend might be attributed to progressively higher proportions of associations of marginal strength which were reinforced by rehearsal. In view of the small and unreliable differences between the conditions of testing it is possible to conclude, however, that the interpolation of MMFR does not introduce a serious bias in the assessment of matching performance.

Finally, the question should be considered of whether the variations in RI may be due, at least in part, to differences in the amount of responding required on the tests of retention. This possibility applies especially to the MMFR test in which *S* was required to recall all the responses associated with a given stimulus. The larger the number of interpolated lists, the more different responses were likely to intervene between successive List 1 responses. Such intervening recalls are a potential source of interference, e.g., they might provide additional opportunities for unlearning or disrupt associations among the responses within a list. There are several findings which make it unlikely, however, that interferences developed during the test played a significant part in determining the observed differences in RI. (*a*) There is no evidence for differential forgetting of items from the final interpolated list which would be subject to the same interferences during the test. In fact, the gain between the last trial of IL and MMFR increases with the number of interpolated lists, as would be expected since the amount of improvement on the final trial should vary inversely with the level of performance. (*b*) There are no progressive increases in the frequencies of misplaced and unassigned responses as a function of the number of lists. (*c*) The trends in RI for the Matching-Only groups parallel those obtained in MMFR. The responses to be matched were segregated according to list. The arrangement of items and the method of recording the matches were such as to lead *S* to proceed from one list of responses to the next. Nevertheless, the pattern of differences in matching is the same as in recall. Thus, it remains likely that the amounts of RI were determined primarily by the conditions of IL.

REFERENCES

ARCHER, E. J. A re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychol. Monogr.*, 1960, **74** (10, Whole No. 497).

BARNES, J. M., & UNDERWOOD, B. J. "Fate" of first-list associations in transfer theory. *J. exp. Psychol.*, 1959, **58**, 97–105.

BRIGGS, G. E. Acquisition, extinction and recovery functions in retroactive inhibition. *J. exp. Psychol.*, 1954, **47**, 285–293.

BRIGGS, G. E. Retroactive inhibition as a function of degree of original and interpolated learning. *J. exp. Psychol.*, 1957, **53**, 60–67.

HAAGEN, C. H. Synonymity, vividness, familiarity, and association value ratings of 400 pairs of common adjectives. *J. Psychol.*, 1949, **27**, 453–463.

HILGARD, E. R. Methods and procedures in the study of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 517–567.

McGOVERN, J. B. Extinction of associations in four transfer paradigms. *Psychol. Monogr.*, 1964, **78** (16, Whole No. 593).

MELTON, A. W., & IRWIN, J. McQ. The influence of degree of interpolated learning on retroactive inhibition and the overt transfer of specific responses. *Amer. J. Psychol.*, 1940, **53**, 173–203.

POSTMAN, L., KEPPEL, G., & STARK, K. Unlearning as a function of the relationship between successive response classes. *J. exp. Psychol.*, 1965, **69**, 111–118.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, 1944.

THUNE, L. E. Warm-up effect as a function of level of practice in verbal learning. *J. exp. Psychol.*, 1951, **42**, 250–256.

THUNE, L. E., & UNDERWOOD, B. J. Retroactive inhibition as a function of degree of interpolated learning. *J. exp. Psychol.*, 1943, **32**, 185–200.

UNDERWOOD, B. J. The effect of successive interpolations on retroactive and proactive inhibition. *Psychol. Monogr.*, 1945, **59** (3, Whole No. 273).

UNDERWOOD, B. J. 'Spontaneous' recovery of verbal associations. *J. exp. Psychol.*, 1948 **38**, 429–439.

UNDERWOOD, B. J. Proactive inhibition as a function of time and degree of prior learning. *J. exp. Psychol.*, 1949, **39**, 24–34.

(Received May 21, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 3, 246–251

ECHO-DETECTION ABILITY OF THE BLIND: SIZE AND DISTANCE FACTORS[1]

CHARLES E. RICE, STEPHEN H. FEINSTEIN, AND RONALD J. SCHUSTERMAN

Stanford Research Institute, Menlo Park, California

The ability of 5 blind Ss to detect metal discs placed in front of them by use of echoes was measured. S was instructed and trained to respond to the presence or absence of these targets after uttering any sound of his choice. Response thresholds were obtained for various size targets at distances ranging from 24 to 108 in. As distance increased, threshold target size increased. The mean auditory angle subtended by a target calculated to be at threshold was 4.63° with an SD of .21°. These data provide a basis for comparing performance using a vocal echo signal with performance using signal characteristics as independent variables.

In recent years the works of Griffin (1958), Kellogg (1961), and Rosenzweig, Riley, and Krech (1955), have all demonstrated the ability of nonhuman species to use auditory information, particularly echoes, in navigational and food-seeking activities. It has also been learned through the work of Supa, Cotzin, and Dallenbach (1944), and others, that both blind and normal people can use echoes effectively to avoid obstacles which loom in their path. The names "facial vision," "echo location," and "echo detection" have been applied to this sonic distance-sensing technique. The great differences in the psychobiology of the bat, porpoise, rat, and human echo detectors makes direct comparison of their abilities difficult, to say the least. Few measurements have been made which quantify the limits of echo-detection ability of any species, though some efforts in this direction have been made with humans by Kohler (1964), Myers and Jones (1958), and Kellogg (1962). Evidence which has been presented to date suggests that the nonhuman species are more skillful than humans and that they are biologically better prepared for obtaining information from echoes.

The physics of sound indicates that the higher the frequency of a sound, the better the resolution of the echoes reflected from small targets or obstacles and, hence, the better the acuity for echo-detection tasks. Since bats and porpoises have the ability to emit and perceive sounds well above the frequency response range of the human auditory apparatus, it seems reasonable to expect that human *S*s, unaided by technical assistance, would have considerably less ability than these nonhumans on echo-detection tasks. The current study is designed to secure a measure of echo-detection ability which represents *S*'s skill using only his natural vocal resources. Specifically, response-threshold measurements were made of the ability of blind *S*s to detect different size stimuli at several distances, as an index of "echo acuity."

METHOD

Subjects.—The *S*s were five blind males, ranging in age from 23 to 30 yr. All *S*s had been blind for at least 5 yr. They were selected on the basis of: (*a*) near total blindness, (*b*) the ability and inclination to travel with relative independence in the course of daily work or recreational activities, (*c*) at least normal intelligence as measured by having made satisfactory progress during their school careers, (*d*) relatively normal audiograms. Three *S*s had no light perception whatever, while the other two could detect very bright light if shone directly into the eye. The audiograms for all but one *S* were within the normal limits of hearing. This *S* (DB) had a "sensorineural loss of mild degree." The *S*s were paid an hourly wage and appeared to be well motivated throughout testing.

---

[1] This investigation was supported in whole by Public Health Service Research Grant NB 04738 from the National Institute of Neurological Diseases and Blindness. The authors are indebted to W. N. Kellogg, who initiated this project.

Apparatus.—The research was carried out in a laboratory designed to provide a near constant auditory environment. The room was 12 × 24 × 9 ft. The walls and ceiling were covered with acoustical tile and the floor with a wall-to-wall carpet. Isolation from external noise was obtained by use of double doors and walls with intermediate air spaces. The modal noise-level measurement was 42 db. on a General Radio 1551C sound-level meter C scale, 32 db. on the B scale, and 26 db. on the A scale. Within the room, 5 ft. from the wall and parallel to the 12-ft. dimension, an 8-ft. wide partition extended from floor to ceiling. It was also covered with acoustical tile. This partition served as the background for stimuli presented to *S*s. The stimuli were circular, .050-in. thick aluminum discs of varied size. A general view of the laboratory room and apparatus is shown in Fig. 1.

The apparatus for presenting the stimuli extended down through the ceiling from a cupola on the roof. A metal track by which the stimuli could be raised and lowered into position extended 30 in. beneath the ceiling. It was located 4 ft. in front of the partition and 6 ft. from either side wall. Stimulus targets were presented to *S* at approximately ear level. This was accomplished by attaching them, by means of a permanent magnet, to a ¼-in. square rod, and lowering the rod along the track to the floor. An apex of this rod was toward *S*. A quiet Slo-syn motor was used to move the stimuli up and down the track. The *S* was seated in a testing chair before the partition and at the appropriate distance from a target. The chair legs were placed in a metal channel, thus assuring the proper orientation to the target. An adjustable headrest on the chair was equipped with a microswitch and buzzer that signaled *S* when his head moved too far forward. In order to control further for auditory and visual cues, *S*s were blindfolded and a white noise source located above *S*'s head was activated between trials.

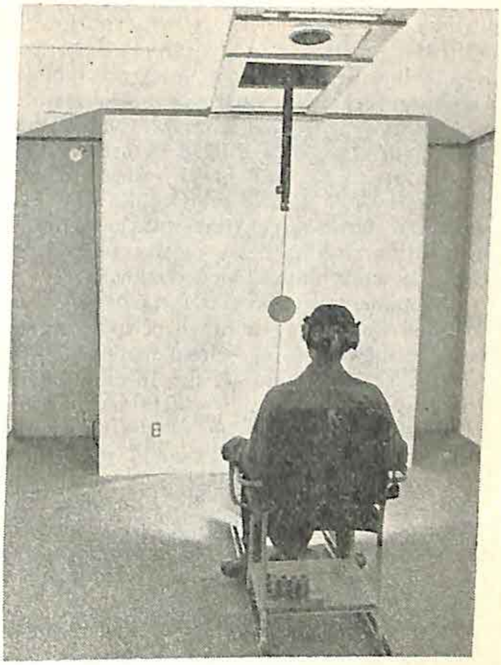


FIG. 1. View of *S* and apparatus with target in position for judgment.

Procedure.—The training of *S*s was divided into two parts: a pretraining procedure and a warm-up period designed to maintain reliable performance.

Pretraining began during the initial visit to the laboratory. After examining the laboratory, *S* was seated in the testing chair and standardized instructions were read. The *S* was told to begin his chosen "noise" after the "ready" signal and then to make his judgment as to presence or absence of the target on future trials.

The *S*'s chair was then adjusted so that there was a distance of 18 in. from his ears to the target. A target, 19.4 in. in diameter, was positioned in front of *S*. He was then required to utter some sound on the signal "ready" and to make a judgment of "yes" or "no" as to presence or absence of a target. This sound was to be any vocalization of *S*'s choice which he felt enhanced his ability to detect the targets. A series of practice trials was run, using an equal number of target and no-target trials in a counterbalanced order. The *S* was given immediate feedback concerning the accuracy of his judgment. When *S* could successfully discriminate between target and no target, a new, smaller target was introduced. This procedure was continued until the smallest target estimated to be detectable with 90–100% accuracy was found. This target was then used as the *largest* in a series of five subsequent stimuli. Each of the four smaller targets was 60% by area of the next larger. A randomized series of 100 presentations of each of the five discs in this series and 100 no-target trials was then judged by *S* without feedback; the results constituted the echo-acuity data at that distance.

A similar, but abbreviated, procedure was followed in locating a range of five targets suitable for measuring echo-detection ability at 24, 30, 36, 42, and 48 in. In addition to these five distances, it was felt desirable to make some measurements of the maximum distance at which targets could readily be discerned. Accordingly, three measurements were made beyond 48 in., extending the distance to a maximum, allowable by the apparatus. The data obtained from the measurements at 24–48 in. suggested that the auditory angle subtended by a target was related to the probability of a "yes" response at any distance from *S*. Using this hypothesis, targets predicted to have high, medium, and low probabilities of detection were made for each of the longer distances: 67, 87, and 108 in.

It was eventually found to be necessary to delete data obtained at the 18-in. distance because some *S*s were detecting the $\frac{1}{4}$-in. rod being used to suspend the target in front of them. Tests in which a target and rod, no target or rod, and rod alone, were presented demonstrated this problem. Similar tests at the 24-in. distance indicated that the rod no longer influenced judgments.

Each of the *S*s were given 5–10 min. of warm-up time before each experimental session in all phases of the experiment. In this practice period, immediate feedback was given by *E* as *S*'s judgment warranted and, therefore, constituted additional training. The *S* could stop a session at any point if he became fatigued, ill, or confused. The warm-up session also served to stabilize *S*'s decision criterion and held false-positive judgments below 20%. By setting such a criterion, it was hoped to obtain comparable measures for all *S*s.

A short series of 30 trials was run in which no noise or echo signal was uttered by *S*, but presence or absence judgments were required. This series was used to determine whether ambient noise, odor, or heat cues were affecting *S*'s response criterion. The *S*s were unable to detect the targets without their echo signals.

RESULTS AND DISCUSSION

The amount of training necessary to progress from the initial 19.4-in. diameter target to the range of five targets used for the first response-threshold measurements varied among *S*s. Once the asymptote of training had been reached, *S*s maintained a consistent level of performance. The time elapsed from the "ready" signal, through the uttering of *S*'s sound, to his response, was normally under 10 sec. The time between judgment of one trial and the next "ready" signal was 10–15 sec.

The echo-ranging noises used by *S*s were all different. The signal produced by CB was a harsh "F" sound which was pulsed several times and followed by "hello, hello." DB created a tongue click, "tsk, tsk, tsk," followed by an extended "sssixes, sseven, eight." DD made an intense tongue click which was sharp and repetitive, originating from the roof of his mouth. JP uttered a variety of sounds, including tongue clicks, lip smacks, and throat clearing. WG used a "hiss-hiss" which was elongated up to 5 sec., and then repeated. In each case, *S* turned his head spontaneously from side to side, sending the signal back and forth across and beyond the targets in a sort of auditory scanning similar to that reported by Kellogg (1962). Once a pattern of noise was established for use as an echo signal, *S* used it consistently throughout the experiments.

Response thresholds.—Response curves in Fig. 2 show the relationship between target size and percentage of "yes" responses for 100 presentations of each target at a given distance. These results show that echo acuity by blind *S*s is a function of both target size and the distance between *S* and target; i.e., the greater the distance, the larger the target must be in order to be detected.

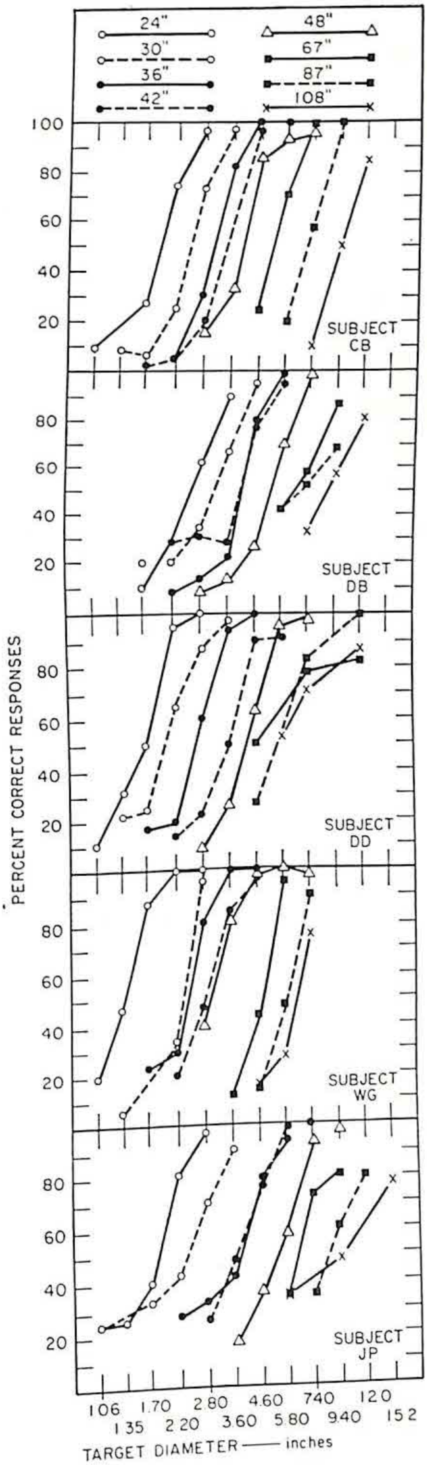
Between 24 in. and 48 in., inclusive, it will be noted that the target-size range used is virtually the same for all *S*s. Beyond 48 in. the targets were not necessarily proportional to one another and only three data points were obtained at each distance. In any case, there is a consistent overlap of target sizes used with *S*s at all distances. This indicates that the echo acuity of this group is rather homogeneous, with no markedly divergent level of skill shown by any of the *S*s.

By comparing response curves for all *S*s at a given distance, their performances may be ranked. Arranging them in this manner for all distances through 48 in. reveals that *S*s tend to remain in the same relative position within the group. (Coefficient of concordance = .78, $p < .01$.)

Within the range of distances considered (24–48 in.), *S*s clearly ranked themselves as follows: WG, DD, CB, JP, and DB, respectively, with DB requiring the largest target at threshold.

Using a linear interpolation, a target diameter that would theoretically yield a "yes" response 50% of the time was designated as the response threshold. These interpolated thresholds are shown in Table 1, along with the mean false-positive level, for each *S*, for the group at each distance, and the overall mean and standard deviation of the false positives. The false-positive data indicate that the desired uniform response criterion was obtained for all *S*s. Thus, the response-threshold comparisons were not unduly affected by differences in the false-positive level.

Maximum distance.—The maximum distance at which targets could be detected was not found. The laboratory conditions in the experiment



[FIG. 2. Percentage of "yes" responses by *S*s to targets of varied diameter at each distance.

TABLE 1

INTERPOLATED THRESHOLD DIAMETERS AND MEAN FALSE-POSITIVE RATES

Distance (In.)	Threshold Diameter (In.)					Mean No. False Positives[a] at Each Distance
	WG	DD	CB	JP	DB	
24	1.4	1.7	2.0	1.8	2.6	10
30	2.4	2.0	2.5	2.4	4.0	9
36	2.5	2.7	3.1	3.8	4.1	6
42	2.9	3.5	3.4	3.7	4.1	6
48	3.0	4.3	4.0	5.4	5.3	4
67	4.7	4.6	5.2	6.5	6.9	10
87	6.0	5.6	6.6	8.6	6.7	9
108	6.5	6.2	9.5	9.8	8.9	8
Mean No. False Positives for each *S*	6	7	4	13	9	

[a] Group mean and *SD* are 7.8 and 2.9, respectively.

prevented moving more than 108 in. back from the targets. As will be noted in the figures and tables presented, no failure in ability to detect the targets occurred. Although fewer data points were used and the proportional relationship of size between the targets was not maintained, the threshold targets are proportionate in size to those at the closer distances. This may be seen in Fig. 3, which combines the data for all *S*s at all distances in group curves. There is less overlap of disc sizes between *S*s beyond 48 in.; hence, mean diameters at the longer distances are based on fewer judgments.

As measurements were obtained, it became apparent that the auditory angles subtended by the threshold targets for all *S*s showed relatively little variability. This observation

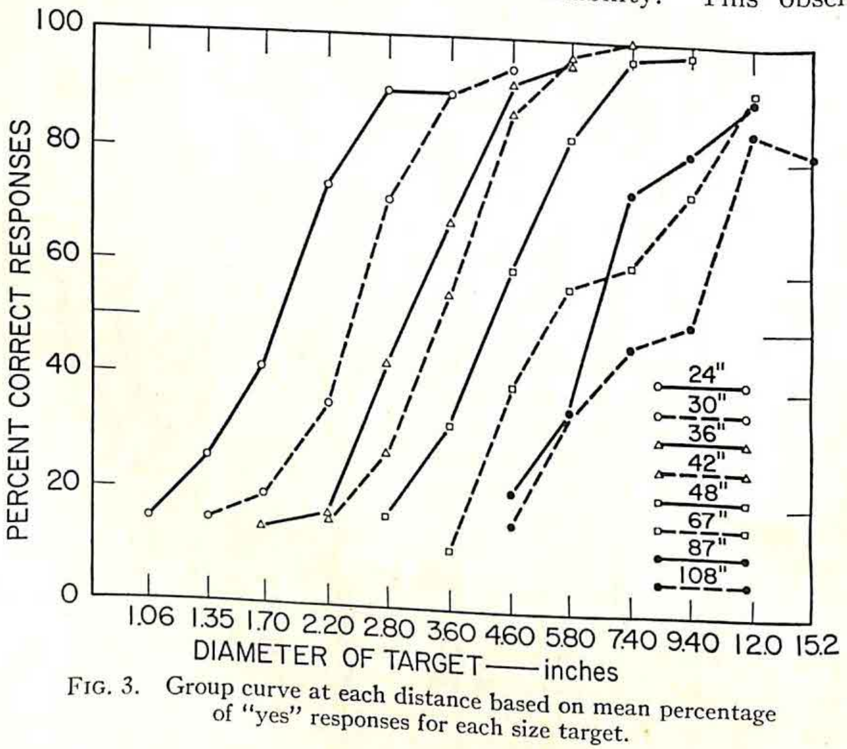


FIG. 3. Group curve at each distance based on mean percentage of "yes" responses for each size target.

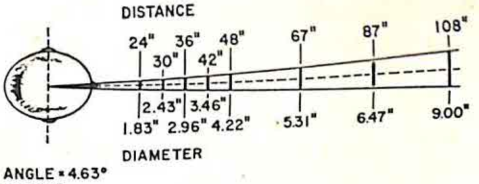


FIG. 4. Group threshold targets at all distances fitted to the mean auditory angle.

lead to the hypothesis that a mean auditory angle for the group might provide meaningful prediction of threshold targets at any distance. This hypothesis was not checked empirically, but post-hoc analysis of the data provide some confirmation.

An interpolated-group threshold target was found for each distance and its auditory angle was calculated. The mean auditory angle for all distances and *S*s is 4.63° with an *SD* of .21°. This angle has been drawn to scale and the group threshold targets were fitted to it in Fig. 4. It appears from these results that, for this group, the auditory angle subtended by the target in front of *S* is a good predictor of *S*'s response threshold at any distance through 108 in. Kellogg (1962) does not specifically state this relationship of auditory angle to size and distance perception in his study, but does present data consistent with our formulation.

Since the auditory angle is regarded as an arc on the horizontal plane at the height of the ears, then it seems likely that width may have been an important factor in detection of the targets. The above hypothesis is consistent with the lateral head movements used by all *S*s to "find the edges of the target." Certainly there is some interaction of width, height, and total area required for sufficient echo information to be obtained.

To our knowledge, no comparable studies of threshold echo discrimination have been reported in the literature for nonhuman species. A bat, however, has been observed (Griffin, 1959) to have a gnat as small as .0002 gm. in its mouth. The resolving power suggested by this observation serves to emphasize the contrast between the natural ability of human and nonhuman skills for this type of task. The results reported here, however, clearly show that human *S*s can gain a measurable amount of information from reflected sound sources. It must be pointed out that generalizations concerning the degree of echo acuity possessed by any experimental group must be qualified in terms of the conditions under which the measurements were conducted. Many variables may affect both the physical sound waves which carry the echo information and the organism's ability to perceive them. It seems at this point, therefore, of minor importance to look at our *S*s' response thresholds as an end in themselves, or to compare our results with those of investigators who have looked at various aspects of echo detection under widely varied circumstances. The present results, nevertheless, serve as a base line or reference point from which to measure the relative importance of each parameter of echo detection to the skill as a whole. Thus, now that it is known what these *S*s can do with their own signals, we are free to investigate the effect of modifying target dimensions, signal characteristics, etc., on their respective response thresholds.

REFERENCES

GRIFFIN, D. R. *Listening in the dark.* New Haven: Yale Univer. Press, 1958.

GRIFFIN, D. R. *Echoes of bats and men.* Garden City, N. Y.: Doubleday, 1959.

KELLOGG, W. N. *Porpoises and sonar.* Chicago: Univer. Chicago Press, 1961.

KELLOGG, W. N. Sonar system of the blind. *Science*, 1962, **137**, 399–404.

KOHLER, J. Orientation by aural clues. *Res. Bull. Amer. Found. Blind*, 1964, **4**, 14–53.

MYERS, S. O., & JONES, C. G. E. F. Obstacle experiments: Second report. *Teacher Blind*, 1958, **46**, 47–62.

ROSENZWEIG, M. R., RILEY, D. A., & KRECH, D. Evidence for echolocation in the rat. *Science*, 1955, **121**, 600.

SUPA, M., COTZIN, M., & DALLENBACH, K. M. "Facial vision," the perception of obstacles by the blind. *Amer. J. Psychol.*, 1944, **57**, 133–183.

(Early publication received
December 31, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 3, 252–263

TASK PREDICTABILITY IN THE ORGANIZATION, ACQUISITION, AND RETENTION OF TRACKING SKILL[1]

DON TRUMBO, MERRILL NOBLE, KENNETH CROSS,[2] AND LYNN ULRICH

Kansas State University

250 male students were assigned to 4 conditions of task predictability, 3 retention intervals, and 2 levels of training in a 4 × 3 × 2 design. Predictability was determined by irregular step-function tasks that differed in the proportions of systematically repeating (predictable) targets, ranging from fixed to random sequences. Integrated absolute error served as a performance criterion. In addition, 6 indexes of temporal-spatial patterning were obtained. Results showed greatest improvement and greatest absolute retention losses for the fixed task and a fixed-direction task added to the design. Intermediately predictable tasks did not differ in error from the random task; however, differences in response organization were found among all tasks. Results suggest the nature of changes in response organization, and indicate that timing may be most crucial for acquiring and maintaining skill.

The purpose of this study was to evaluate the effects of task predictability on the temporal-spatial organization, the acquisition and the retention of tracking skill. Task predictability, as used in this study, is a task variable, not unlike task coherency or redundancy, whereby the kind and amount of advance information available to the operator is systematically varied. The amount of advance information determines the extent to which perceptual anticipation is possible, as Poulton (1950, 1952, 1957a) has pointed out. The distinction between response *organization* and *acquisition* signifies our interest in the temporal-spatial patterns of behavior as well as in the outcome criteria of performance. By *response organization* we refer to the patterns or strategies of response which *S* develops in the various task conditions, whereas *task organization* refers to the predictability or coherency of the input signal.

The significance of task-organization variables for the understanding of skilled performance has been acknowledged for some time by both British and American investigators. Poulton (1950, 1952, 1957a, 1957b) was among the first to recognize the significance of this class of variables for perceptual anticipation in skill tasks. Thus, he distinguished anticipation resulting from *S*'s learning of the predictable statistical or informational properties of the input from anticipation ("receptor anticipation") made possible by actual preview of the signal course. The latter has also been studied by Poulton (1954),

---

[1] This research was supported by the Air Force Office of Scientific Research under Grant No. AF-AFOSR 62-17. The authors are indebted to Lowell Schipper for his contributions to the design of this study. William Morton and Theodore Evans, participants in a National Science Foundation Undergraduate Research Participation Program conducted by the Department of Psychology, Kansas State University, assisted in data collection and analysis.

[2] Now at the Decision Sciences Laboratory, Hanscom Field, Bedford, Massachusetts.

Leonard (1953), and Crossman (1960), among others.

In the United States, Briggs and Naylor (1962), Naylor and Briggs (1963b), and Adams and Webber (1963) have shown continued interest in this class of variables. For example, Naylor and Briggs (1963b) have distinguished task complexity (defined in terms of digram probabilities within a single task dimension) from task organization (defined in terms of interdependencies among different task dimensions) and have shown the significance of both types of variables for skilled performance and acquisition.

While there has been a good deal of interest in task-organization variables, two limitations in the research are apparent. First, the experimental evidence is predominately for tasks involving continuous input signals, such as tracking tasks employing more or less complex combinations of sinusoids, (e.g., Hartman & Fitts, 1955) and it may be that other variables inherent to such tasks interact with task organization. What evidence there is for the effects of organizational variables with discrete signals, such as a step-function tracking task, does not conform closely to the prediction that ". . . progressive increases in redundancy should result in progressive improvement in skilled performance . . . . [Fitts, Noble, Bahrick, & Briggs, 1959, pp. 8–14]" For example, Bahrick and Shelly (1958) failed to find any important difference in performance on four response-key tasks which varied from 100% to 0% in redundancy, although performance on a time-shared task was directly related to redundancy in the primary task, supporting indirectly the prediction.

The second limitation in the literature is a paucity of evidence on the qualitative changes in response patterns as a result of practice with tasks of various degrees of organization. While it is widely held that an understanding of skill must come from investigation of the continuous time-varying response functions, researchers continue for the most part to deal with outcome indexes, such as time-on-target or integrated error scores. It is our contention that investigation of conditions of task organization on the one hand and detailed analysis of the consequent response organization on the other is a fruitful approach to skilled performance. Such an attack should allow us to explore the intervening information processing and response strategies which appear to be of key importance to the development of skill.

Turning to the question of long-term retention of learned skill, for the main part task-organization variables have not been studied (Naylor & Briggs, 1961). Rather, researchers have been concerned with amount of practice (Ammons, Farr, Bloch, Neumann, Dey, Marion, & Ammons, 1958; Duncan & Underwood, 1953; Hammerton, 1963), length of the retention interval (Ammons et al., 1958; Duncan & Underwood, 1953; Jahnke & Duncan, 1956; Neumann & Ammons, 1957; Van Dusen & Schlosberg, 1948) and distribution of practice (Jahnke & Duncan, 1956). A recent exception is to be found in Naylor, Briggs, and Reed (1962), a study involving 1- and 4-wk. retention intervals. These investigators varied the organization—defined in terms of redundancy—not of the primary (tracking) task, but of a secondary procedural task. The organization variable was found to be related to the amount of retention for both the procedural and the primary tasks under moderate amounts of training.

The present study incorporated both amount of training and length of retention interval together with the organization variable, task predictability, as independent variables. The analyses of the learning and retention data included overall performance, as reflected in integrated error scores, and, for selected conditions, analytic information about temporal and spatial accuracy from six indexes taken from continuous records of performance.

METHOD

Subjects.—A total of 250 right-handed male students, between 17 and 26 yr. of age, were assigned, 10 each, to 25 experimental conditions.[3] The *S*s were paid for their services.

Apparatus.—The apparatus was the Kansas State University Versatile Electronic Tracking Apparatus (VETA) which has been described elsewhere in detail (Trumbo, Eslinger, Noble, & Cross, 1963).

Programs are punched on paper tape, read out by a commercial tape reader, converted to analog voltages by a digital-to-analog converter and a flip-flop circuit, and displayed on a CRT as a $\frac{1}{2}$-in. long vertical hairline. By moving a lateral arm control, pivoted at the elbow, *S* controls a similar vertical line—the cursor. The cursor is displayed below the target on the CRT and overlaps it by $\frac{1}{8}$ in. Scoring is accomplished by an operational amplifier manifold which obtains the absolute difference between target and cursor (momentary absolute error), and integrates this difference over the trial (integrated absolute error). Input, output, and momentary error are simultaneously recorded, as desired, on magnetic tape, while integrated error is recorded by *E* on each trial from a voltmeter.

Tasks.—Each of the 10 groups received irregular step-function inputs at one target per sec. (.5 cps). Targets could appear at any 1 of 15 equidistant positions on the horizontal axis of the CRT. Trials were 60 sec., separated by 24 sec., and anticipated by a 2-sec. warning buzzer.

Experimental conditions.—The experimental design was basically a 4 × 3 × 2 factorial with four conditions of task predictability, three retention intervals, and two levels of training. A fifth condition of task predictability was included, but only under the longest retention and training conditions (see Task S-P below).

Task predictability.—This variable was defined in terms of the proportion of repeating targets in a sequence. For the Predictable Task (Task P), 12 targets selected randomly from the 15 possible target positions, were repeated in the same order five times per trial and throughout all trials. Coding the target positions from left (1) to right (15) this sequence was: 4, 1, 11, 10, 4, 7, 1, 15, 5, 4, 5, 10. For the unpredictable Random Task (Task R), a sequence 240 targets long was selected randomly, except that a target could not be immediately repeated.

Two conditions of intermediate predictability were based on 12-unit sequences. In Task I-1, every *second* target was selected anew (randomly, as above) on each repetition of the sequence, but the remaining six targets were repeated in order, throughout training. Similarly, in Task I-2 every *third* target was selected anew on each repetition. Thus, the fixed targets were repeated five times per trial for both tasks.

The "extra" condition was labeled semipredictable (Task S-P). It was selected as in Task R, but with a simple, thus predictable, LRLR directional pattern.

Training levels.—All *S*s were given 7 60-sec. trials on a common random pretraining task, followed on Day 1 by 20 training trials. Low Training *S*s received 30 trials on Day 2, for a total of 50 trials. High Training *S*s were given 30 and 20 trials on Days 3 and 4, or 100 trials in all.

Retention intervals.—For each combination of task and training level, one third of the *S*s were assigned to each of three retention intervals: one wk., 1 mo., and 5 mo. At recall, all *S*s received 15 trials on their training tasks.

Procedures.—The *S*s were assigned randomly to the 25 conditions and to the training schedule, except that 5-mo. retention *S*s were trained first to assure their availability for retention.

After a common general description of the study and the task, each *S* was given the seven pretraining trials. Then common detailed instructions were given, including illustrations of various sources of error such as lags, leads, overshoots, and slow primary movements. The *S*s were instructed to look for patterns but were not told the nature of their own pattern. On successive days *S*s were told only that the task remained the same.

---

[3] Actually, 265 *S*s received some training. A total of 10 *S*s who failed to complete the training phase or the 1-wk. and 1-mo. retention conditions were replaced. Eleven *S*s were trained in each 5-mo. retention-interval condition to cover losses; however, three of these nine groups had only 9 *S*s while two groups had 11 *S*s during retention tests.

Extrinsic feedback was provided to reduce inter-*S* variability and to discourage inefficient response patterns. For each block of five trials, *S*s were told via an intercom that their error scores were "OK" or "worse" than their scores for the previous block.

Performance measures.—The overall performance criterion was absolute error integrated over individual trials. In addition, three indexes of temporal accuracy and three of spatial accuracy were obtained from oscillographic records for eight training and three retention trials (Trials 1, 20, 21, 49, 51, 79, 81, 99, and Retention Trials 1, 2, 3). These scores were obtained only for the 50 *S*s with high training and the 5-mo. retention interval. Temporal indexes were *mean lag time*, *mean lead time*, each scored to the nearest 50 msec. and including all instances of zero lead or lag, and *beneficial anticipations*, scored whenever leads and lags were less than 150 msec. Number of *overshoots*, *undershoots*, and *incorrect anticipations* were scored as indexes of spatial accuracy. Overshoots and undershoots were scored whenever the primary movement missed the target by the equivalent of 5 mm. or more on the CRT. This was approximately two thirds the distance between adjacent target locations.

RESULTS

Figures 1 and 2 summarize the integrated error data for the various

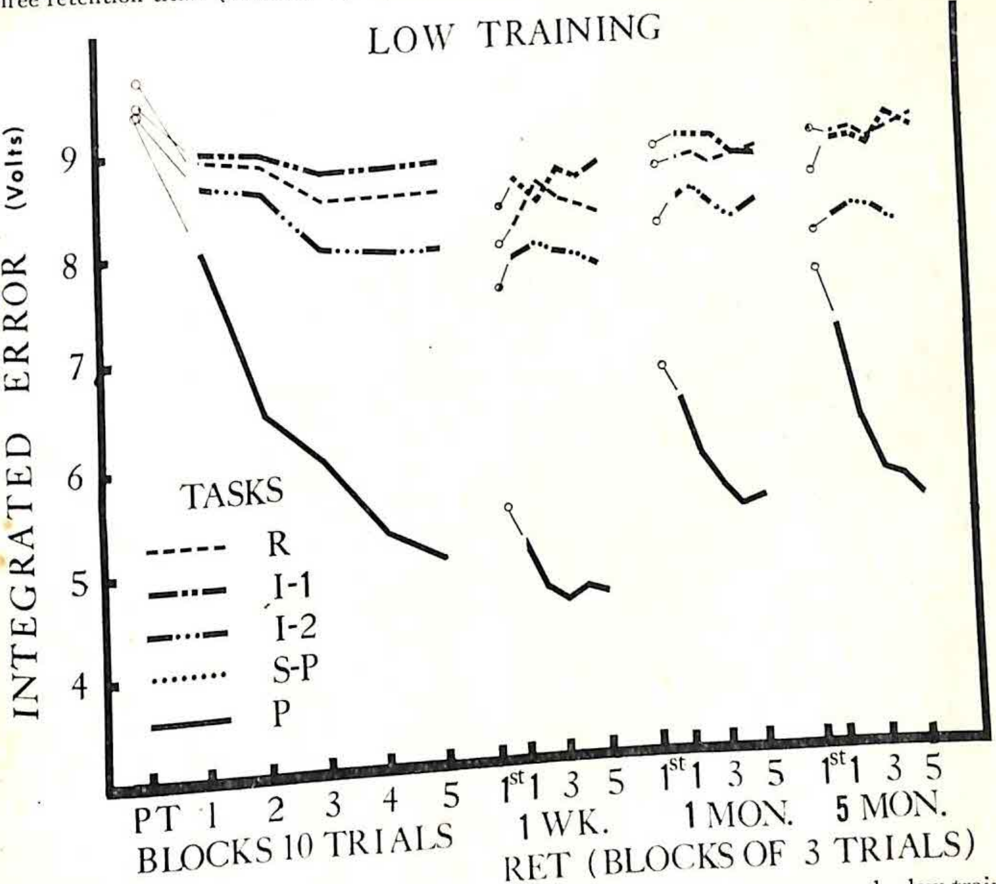


FIG. 1. Integrated error (in arbitrary voltage units) for all task conditions under low training. (Data points are for blocks of 10 trials with 30 *S*s in training and for blocks of 3 trials with 10 *S*s in retention. Open circles at the left represent means of 7 pretraining trials. At the right, these circles indicate initial retention trial scores.)

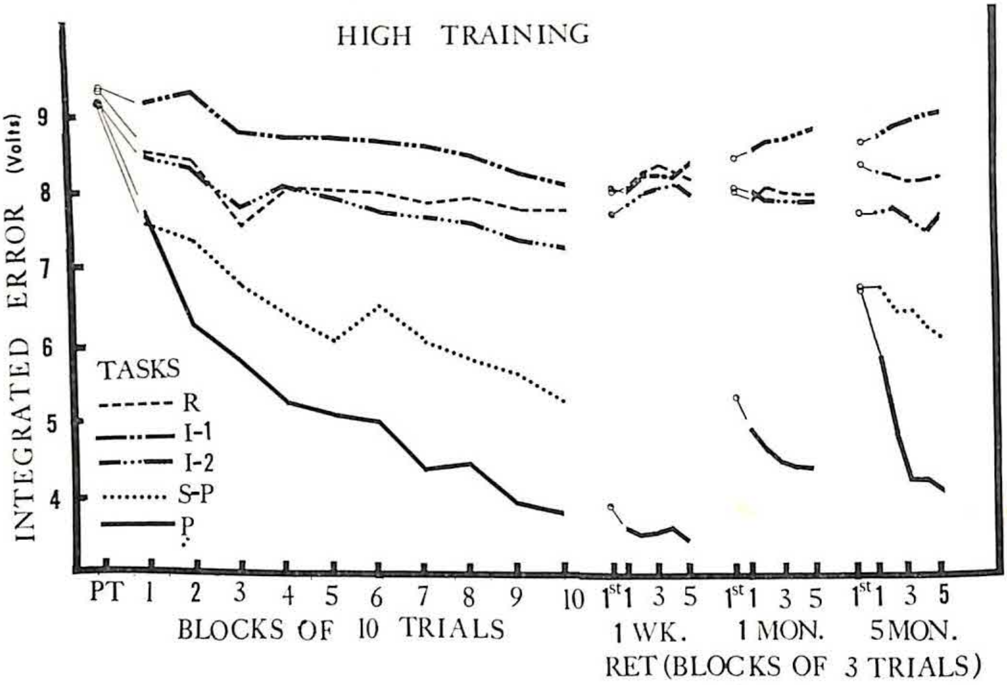


FIG. 2. Integrated error data for all task conditions under high training. (The figure is keyed the same as Fig. 1, except Group S-P—dotted line—has been added.)

conditions during both training and recall sessions. Figure 1 shows the acquisition data in blocks of 10 trials and retention data in blocks of 3 trials for *S*s with low training. Figure 2 presents these data for high training including the extra S-P conditions. Curves for the original training sessions are each based on 30 *S*s who were then divided into the three retention interval subgroups of 10 *S*s each. While this precludes a direct comparison of training and retention performance, these subgroups were not found to differ appreciably during training.

At the end of Low Training, Group P had significantly less error than all other groups, while the difference between Groups I-2 and R barely reached the .05 level. After High Training, Groups P and S-P had reliably less error than the remaining groups, but Group I-2 was no longer significantly different from Group R. Tests of differences were based on means of the last five trials (*t* tests with $p < .05$).

It is apparent that retention losses were correlated with length of the retention interval and that performance at recall was superior for the High Training conditions. In addition, absolute losses over the retention periods were greater for the more predictable patterns; but then, they had much greater gains during training, and hence more to lose.

There appears to be no completely satisfactory way to compare the relative losses of groups which had attained different levels of performance during training. No statistical analysis or transformation which we considered yields results that are entirely satisfactory. Some information is gained, however, by comparing the retention losses relative to the total improvement of each group. Total improvement was assessed as the difference between the means for pretraining and for the last block of training trials. Since pretraining preceded detailed instructions, it seemed to represent a good lower limit of performance. Losses were calculated as the difference between means for the last block of training and for the initial three retention trials.

While the reliability of these data was not investigated, the results suggested that relative losses were greatest for Task P under low but not under high training conditions. The *S*s on the random task showed gains on the first recall trial after 1 wk. and 1 mo. with low training and after 1 wk. with high training. The low training I-2 *S*s showed a similar gain after 1-wk. retention. These gains suggested either the dissipation of interfering response sets, the relatively large benefits of additional practice afforded by the recall trial, or both.

With the exception of the I-1 group, all groups showed some evidence of relearning during the 15 retention trials. Why the I-1 group showed a loss during these trials was far from apparent.

The benefits of additional training were particularly apparent in the retention trials for Group P. With low training, only 1-wk. retention *S*s regained their original level during the 15 relearning trials, whereas with high training even the 5-mo. *S*s showed essentially complete relearning. Finally, although Task P *S*s showed less relative loss with high than with low training, *S*s on Tasks R and I-2 had greater relative losses after high training. This suggested that the overlearning of the predictable pattern offset the greater loss potential for Group P, but for *S*s on the less predictable patterns the effect was primarily to increase the loss potential without overlearning benefits.

Analytic scores.—Data for the six analytic scores were taken from oscillographic records of the 50 high-training, 5-mo. retention *S*s. These data are presented in Fig. 3 and 4, for the eight selected learning trials and for Trials 1, 2, and 3 of retention. Additional comparisons were made between the best three and poorest three *S*s in each group, using integrated error for the last block of training trials as a criterion. These latter data are not presented graphically, but the results are discussed where they help to clarify other findings.

Temporal accuracy.—Figures 3A and 3B present the mean lag and mean lead times, respectively, for eight training and three retention trials. Figure 3C shows the percentage of beneficial anticipations for the same trials. The training trials were selected, one early and one late in each of the four sessions, to reflect differences within and between sessions as well as throughout training.

Lags.—Each of the five groups decreased mean lag time over the pretraining period with the more predictable groups (P and S-P) showing the greatest change. The latter two groups differed significantly ($p < .05$) from the remaining three groups which did not differ one from another. Comparison of Day 1 (Trials 1 and 20) and Day 4 (Trials 81 and 99) means revealed that all groups except I-1 had significant decreases in lag time ($p < .05$).

Comparison of lag scores on the first recall trial with the first trial of Day 4 (Trial 81) showed greater lag times at retention for all except the I-2 group. Mean values at recall were approximately equal to those for Day 2 of training (Trial 21).

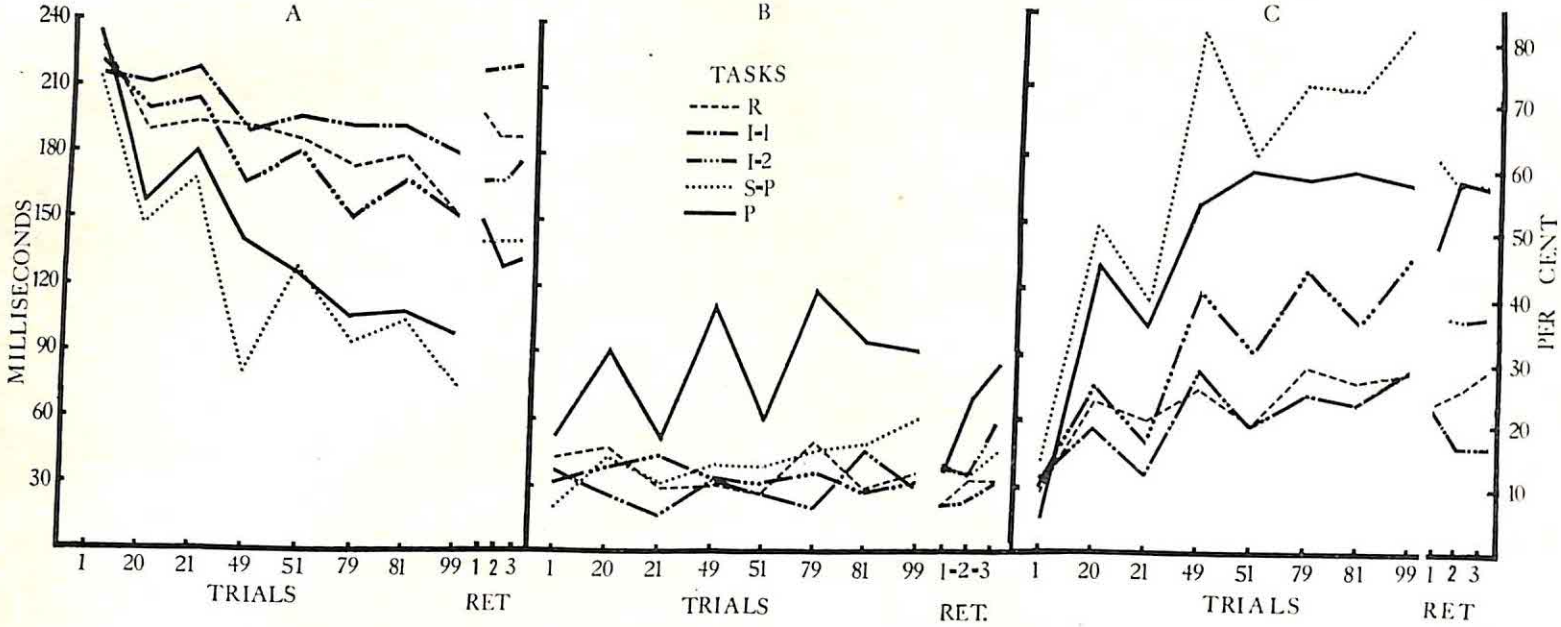


FIG. 3. Data for timing accuracy, including (A) mean lag scores, (B) mean lead scores (in milliseconds), and (C) mean beneficial anticipations (in percentage of total targets).

Furthermore, the *number* of lags increased over the retention interval to a frequency comparable to that of early training.

Within-group comparisons indicated that best *S*s had markedly fewer lags and shorter mean lag times than the poorest *S*s in each task condition. Best *S*s tended to reduce both frequency and lag time on all tasks whereas poorest *S*s showed little change except on Tasks P and S-P, the most predictable conditions. Retention losses were comparable for best and poorest *S*s, but despite the losses, best *S*s had shorter lags at retention than poorest *S*s had at the end of training.

Leads.—Only *S*s on Tasks P and S-P showed any consistent trends in mean lead time during training (Fig. 3B). Group S-P had a significant increase from Day 1 to Day 4 and Group P had a significantly higher mean than the remaining groups at the end of training ($p < .05$). As with the lag data, there were pronounced losses between sessions, particularly for Group P, indicating that timing may be highly susceptible to forgetting, especially with modest training.

This conclusion was supported by the recall data, wherein all groups showed complete losses of their original gains in lead time, with, however, evidence of rapid recovery during the first three relearning trials.

Within the groups, the best *S*s had greater increases both in frequency of leads and mean lead times than poorest *S*s, except for the I-1 and I-2 tasks. In the latter groups, even the best *S*s had markedly fewer leads and lower lead times than the other groups, including Group R. For example, on the I-2 task, best *S*s reached a maximum of 43% leads; in contrast with 87% for their counterparts on Task R. Thus, there appears to have been a marked inhibition of leading responses in the intermediate tasks.

Consistent results were found at retention. Best *S*s recovered about 50% of their leads by the end of the third trial, but the frequency remained below 10% in the I-1 and I-2 tasks. Poorest *S*s had very few leads at retention, except on Task P where they reached 20% by Trial 3.

Beneficial anticipations.—Figure 3C indicates that each group increased in percentage of beneficial anticipations during training. Increases from Day 1 to Day 4 were significant for all except the I-1 condition. Increases were proportional to the degree of task predictability, again, with the exception of Group I-1. As with the lag and lead indexes, there is evidence of considerable loss between sessions, indicating that timing is highly susceptible to forgetting with even brief periods of no practice.

However, there is less evidence of loss in retention on this measure than on the lead and lag indexes. Groups P and S-P regressed to about midtraining performance, but the remaining groups showed little or no loss. This may reflect, in part, an insensitivity of the beneficial anticipation measure to changes from lead to lag or in response variability. As a rather gross bandwidth score, it is subject to the general criticisms of time-on-target scores (Bahrick, Fitts, & Briggs, 1957).

Comparisons within groups revealed that the poorest *S*s were able to increase the frequency of beneficial anticipations only on Tasks P and S-P, whereas the best *S*s showed increases on all tasks and a much higher frequency than poorest *S*s on Tasks I-1, I-2, and R.

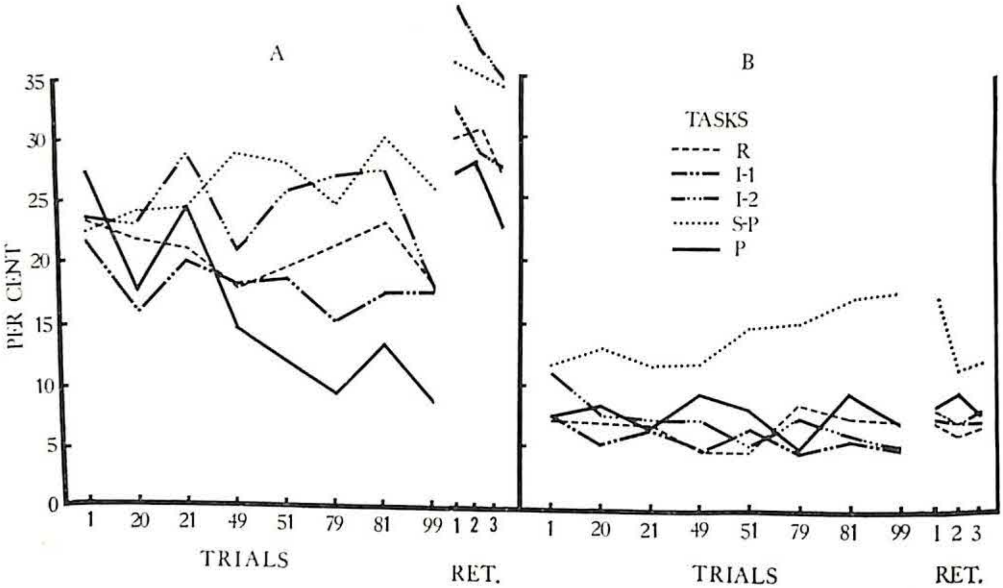


FIG. 4. Data for spatial accuracy, including (A) overshoots and (B) undershoots. (Both indexes are given in percentages of total targets tracked.)

Spatial accuracy.—Figures 4A and 4B present the data for overshoots and undershoots, respectively. Data for incorrect anticipations are not shown, since the low frequency for this type of error made their reliability suspect.

Overshoots.—Only *S*s on Task P significantly reduced the number of overshoots (Fig. 4A). By contrast, the S-P group tended to increase overshoots, the remaining groups showing no consistent change.

At recall, all groups had a marked increase in overshoots with Group P regressing to its Day 1 level and all other groups showing higher frequencies than at any point in training.

Intragroup comparisons revealed that the best *S*s *increased* the number of overshoots in all tasks except Task P where they decreased, while the poorest *S*s had slight decreases on all tasks. At the end of training best *S*s had from 5 to 25% more overshoots for all tasks except Task P where both best and poorest *S*s had reached a low of about 5%. Thus, it appeared that the poorest *S*s (in terms of integrated error) had better spatial accuracy and, furthermore, showed improvement, whereas, except for the fully predictable task, best *S*s became less accurate as training progressed.

Undershoots.—The undershoot data, presented in Fig. 4B, show little change throughout training. However, Group S-P had significantly more undershoots than the other groups at the end of training. Coupled with their tendency to increase overshoots, this gave the S-P group the poorest record for spatial accuracy.

Best *S*s on Tasks P, S-P, and R had more undershoots than poorest *S*s, but there were no differences on the I-1 and I-2 tasks.

Finally, there was little evidence of change in undershoots over the 5-mo. retention interval, either from the total group data or the within-group comparisons.

DISCUSSION

Acquisition.—The most striking finding from the integrated error data was the failure of *S*s on the I-1 and I-2 tasks to achieve a higher level of performance than *S*s with an essentially random task. For these intermediate groups one half to two thirds of the targets were predictable and repeated, in order, 500 times during training. Furthermore, for the I-2 group four sequential pairs of targets received 500 repetitions. Yet there was no evidence to indicate that these predictable elements were learned. Average lead and lag data, scored separately for predictable and unpredictable elements of the task, did not indicate superior anticipation of repeating targets.

Unlike the integrated error data, measures that reflect temporal characteristics of responding, indicated some differences in response patterning among Tasks I-1, I-2, and R. Specifically, it appeared that *S*s presented with a random task were much more apt to anticipate target displacements than *S*s with either of the "I" patterns, despite the fact that the "I" patterns are objectively more predictable, a fact which would be expected to facilitate rather than inhibit anticipatory responding.

In contrast to the results with the "I" groups, the S-P group, for which direction only was predictable, showed substantial gains over the R group in reducing error. Thus, the addition of objective predictability by fixing as many as two thirds of the target positions in a 12-unit sequence has no appreciable affect on the tracking performance with a step-function tracking task, whereas fixing predictability of direction by imposing the simple alternation rule has a highly significant effect. If the position is taken that movements are preprogramed (see Fitts et al., 1959), then it appears that, with the tasks studied here, conditions which permit complete preprograming of directional aspects are more beneficial than those which permit only partial preprograming of both directional and extent aspects. In other words, advance information which leaves zero uncertainty about direction for all movements is more beneficial to performance than advance information which leaves zero uncertainty about both *direction* and *extent* for as many as two thirds of the movements.

Two other findings from the analytic scores suggest tentative conclusions about the response strategies organized to meet the various task conditions. First, the best *S*s in nearly all cases had poorer results on the measures of spatial accuracy than the poorest *S*s. In some cases the better *S*s actually showed decreased spatial accuracy as training progressed. Similarly, the S-P group, which had the second lowest error scores and changes on the temporal indexes comparable to those for Group P, tended to increase overshoots throughout training. Taken together, these results indicate that the more successful *S*s emphasized timing accuracy even at a cost to spatial accuracy. The S-P group, which showed considerable improvement in timing, anticipated target displacements even though they could not predict the amplitudes of the required movements. The best they could do was to anticipate, on the basis of pooled experience with the task, with movements of mean amplitudes. If this were true, they would be expected to undershoot the long displacements and overshoot the short ones, the net result being relatively high frequencies of both types of errors. The data, summarized in Fig. 4A and 4B, support this interpretation. Group S-P led in overshoots and undershoots and tended to increase the frequency of each type of error throughout training. Inspection of oscillographic records indicated that *S*s on Task S-P increased response rate and acceleration, factors which would also result in increased oscillations and spatial inaccuracy. That this strategy "paid off" is apparent from the relatively good criterion performance of the S-P group. It is also apparent from the intragroup comparisons, where, as indicated, best *S*s had better temporal, but poorer spatial scores than poorest *S*s on all but the completely predictable task. It may be noted that the strategy of the best *S*s could have been predicted from servo theory. Specifically, in developing an automatic control device to null system error following a step-function input, the servo engineer would build an underdamped rather than a critically damped system. This would minimize RMS error, but result in increased oscillations, analogous to the poorer spatial accuracy of our best *S*s. Thus, the good *S*s were acting in the best tradition of servo engineering with one innovation; namely, they further reduced their error by anticipating rather than lagging behind target displacements.

Retention.—Contrary to the general tenor of findings on retention of learned skill (Naylor & Briggs, 1961), our results showed substantial losses during the no-practice interval. Absolute losses were greatest for those tasks in which greatest gains were realized during training, and were proportional to the length of the retention interval. Both recall and relearning performance were better for the high-training than for the low-training condition, but even with high training there was considerable forgetting after the 1-mo. and 5-mo. intervals.

After 5 mo. of no practice the initial recall performance of *S*s trained on a completely fixed pattern (Task P) was indistinguishable from that of *S*s trained on a pattern with only direction fixed (Task S-P). This suggests that the additional information about the target course (i.e., a fixed sequence of positions) which resulted in significantly lower error scores for Group P during training was of no advantage in the retention of the skill.

Each of the three indexes of timing accuracy documented some changes during the retention interval. These changes were in the form of decreased lead time and frequency of leads, increased lag time and frequency of lags, and generally poorer timing, as indicated by the decrease of beneficial anticipations. These changes were generally correlated with the degree of task predictability and with differences in overall performance as manifested in the integrated error scores.

The analytic scores, in addition to providing evidence as to the nature of different response strategies developed during training, suggest more clearly than any evidence we have seen specifically what changes occur in the course of forgetting of learned skill. At least with tasks such as those employed here, it appears that temporal rather than spatial accuracy is the more crucial aspect of skill retention and the most susceptible to loss. The extent to which this is true of skilled performance in general must, of course, wait upon further research involving a variety of tasks. Nevertheless, these results are consistent with recent findings of Naylor and Briggs (1963a) which indicate that rehearsal of timing is more effective in maintaining skilled performance than rehearsal of spatial aspects, and those of Adams and Creamer (1962) which show that pretraining on timing provides greater facilitation of skill acquisition than pretraining of other task components.

Finally, the analytic results strongly point toward the role of response strategies adopted by *S* in the development and retention of skill. These strategies appear to be attributable in part to task-organization variables which determine the kind and amount of advance information available to *S*. Individual differences in performance were great, however, and further research is needed to determine the cognitive and other factors involved in intragroup differences in response patterns.

REFERENCES

Adams, J. A., & Creamer, L. R. Anticipatory timing of continuous and discrete responses. *J. exp. Psychol.*, 1962, **63**, 84–90.

Adams, J. A., & Webber, C. E. Monte Carlo model of tracking behavior. *Hum. Factors*, 1963, **5**, 81–102.

Ammons, R. B., Farr, R. G., Bloch, E., Neumann, E., Dey, M., Marion, R., & Ammons, C. H. Long-term retention of perceptual-motor skills, *J. exp. Psychol.*, 1958, **55**, 318–328.

Bahrick, H. P., Fitts, P. M., & Briggs, G. E. Learning curves—Facts or artifacts? *Psychol. Bull.*, 1957, **54**, 256–268.

Bahrick, H. P., & Shelly, C. Time sharing as an index of automatization. *J. exp. Psychol.*, 1958, **56**, 288–293.

Briggs, G. E., & Naylor, J. C. The relative efficiency of several training methods as a function of transfer task complexity. *J. exp. Psychol.*, 1962, **64**, 505–512.

Crossman, E. R. F. W. The information-capacity of the human motor-system in pursuit tracking. *Quart. J. exp. Psychol.*, 1960, **12**, 1–16.

Duncan, C. P., & Underwood, B. J. Retention of transfer in motor learning after twenty-four hours and after fourteen months. *J. exp. Psychol.*, 1953, **46**, 445–452.

Fitts, P. M., Noble, M. E., Bahrick, H. P., & Briggs, G. E. Skilled performance: Parts I and II. *USAF WADC final Rep.*, 1959, No. AF 41 (657)-70.

Hammerton, M. Retention of learning in a difficult tracking task. *J. exp. Psychol.*, 1963, **66**, 108–110.

Hartman, B. P., & Fitts, P. M. Relation of stimulus and response amplitude to tracking performance. *J. exp. Psychol.*, 1955, **49**, 82–92.

Jahnke, J. C., & Duncan, C. P. Reminiscence and forgetting in motor learning after extended rest intervals. *J. exp. Psychol.*, 1956, **52**, 273–282.

Leonard, J. A. Advance information in sensori-motor skills. *Quart. J. exp. Psychol.* 1953, **5**, 141–49.

Naylor, J. C., & Briggs, G. E. Long-term retention of learned skills: A review of the literature. *USAF ASD tech. Rep.*, 1961, No. 61–390.

Naylor, J. C., & Briggs, G. E. Effects of rehearsal of temporal and spatial aspects on the long-term retention of a procedural skill. *J. appl. Psychol.*, 1963, **47**, 120–126. (a)

Naylor, J. C., & Briggs, G. E., Effects of task complexity and task organization on the relative efficiency of part and whole training methods. *J. exp. Psychol.*, 1963, **65**, 217–24. (b)

Naylor, J. C., Briggs, G. E., & Reed, W. G. The effects of task organization, training time, and retention interval on the retention of skill. *USAF AMRL tech. docum. Rep.*, 1962, No. AMRL-TDR-26-107.

Neumann, E., & Ammons, R. B. Acquisition and long-term retention of a simple serial perceptual-motor skill. *J. exp. Psychol.*, 1957, **53**, 159–161.

Poulton, E. C. Perceptual anticipation and reaction time. *Quart. J. exp. Psychol.*, 1950, **2**, 99–112.

Poulton, E. C. The basis of perceptual anticipation in tracking. *Brit. J. Psychol.*, 1952, **43**, 295–302.

Poulton, E. C. Eye-hand span in simple serial tasks. *J. exp. Psychol.*, 1954, **47**, 403–10.

Poulton, E. C. Learning the statistical properties of the input in pursuit tracking. *J. exp. Psychol.*, 1957, **54**, 28–32. (a)

Poulton, E. C. On prediction in skilled performance. *Psychol. Bull.*, 1957, **54**, 467–478. (b)

Trumbo, D., Eslinger, R., Noble, M., & Cross, K. A versatile electronic tracking apparatus (VETA). *Percept. mot. Skills*, 1963, **16**, 649–656.

Van Dusen, F., & Schlosberg, H. Further studies of the retention of verbal and motor skills. *J. exp. Psychol.*, 1948, **38**, 526–534.

(Received May 14, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 3, 264–269

INTERLIST RESPONSE MEANINGFULNESS AND TRANSFER EFFECTS UNDER THE A-B, A-C PARADIGM[1]

L. R. GOULET[2]

Saint Louis University

Paired-associate transfer was studied under the A-B, A-C paradigm with treatments varying in List 1 and List 2 response meaningfulness (M). The levels of response M were either high (H) or low (L) on both lists (H_1-H_2 or L_1-L_2) or conformed to a high-low (H_1-L_2) or low-high (L_1-H_2) relationship. The predictions were made that interference from List 1 (negative transfer) would be greatest under conditions of high List 1 response M (H_1-H_2 and H_1-L_2), but that the interference would be partially offset in treatments involving disparate levels of List 1 and 2 response M (H_1-L_2 and L_1-H_2). The reduction in interference was attributed to an increased opportunity to differentiate the interlist response systems. Both predictions were supported. The results were interpreted within Postman's recent interpretation of interference theory and implications relating response M and transfer were provided.

Two studies (Jung, 1963; Merikle & Battig, 1963) have recently determined the importance of response meaningfulness (M) in affecting the direction and magnitude of transfer in paired-associate (PA) experiments. When List 1 responses are carried over to List 2 (C-B and A-Br paradigms), negative transfer declines and shifts to positive transfer (relative to C-D pairs) as the level of List 1 response M is decreased. When new responses are used on List 2 (A-C paradigm), similar results are obtained except that positive transfer under low levels of response M is not found.

The response M transfer relationship under the C-B (stimuli different, responses identical) and A-Br (stimuli and responses identical but re-paired on List 2) paradigms is adequately accounted for by Underwood and Schulz's (1960) stage analysis of PA learning. This analysis assumes that the difficulty of response integration (learning) is inversely related to response M. The resultant transfer of response learning to List 2, then, is expected to be greater when List 1 is comprised of low M responses (Jung, 1963). However, the response-integration hypothesis cannot account for the analogous transfer function under the A-C paradigm inasmuch as no source of transfer from response learning is possible on List 2.

The present experiment attempted a further analysis of the relationship between response M and transfer under the A-C paradigm. This was accomplished by means of a factorial design involving responses of high (H) and low (L) meaningfulness on the two lists. The H_1-H_2 (responses of high M on both lists) and L_1-L_2 (responses of low M on both lists) treatments replicated those employed by Jung (1963). Jung postulated that the greater negative transfer effect found for H_1-H_2 lists than for L_1-L_2 lists was attributable to interference effects when high M responses comprise List 1. However, the identification of the interference (negative transfer) effect as a function of List 1 response M was not validly permitted in this study in that the factorial combinations of List 1 and List 2 response M were not represented. Jung's (1963) hypothesis is partially supported by the interference conception of PA transfer (e.g., Postman, 1961, 1963) which assumes that negative transfer on List 2 is a function of competition of List 1 (A-B) associations during A-C learning. Exactly why the List 1 interference is greater under high-response M (relative to low-response M) has not been explained, but if Jung's hypothesis is correct, negative transfer on List 2 should be independent of List 2 response M.

---

[1] This paper is based on a dissertation submitted to the Graduate School of Saint Louis University in partial fulfillment of the requirements for the PhD degree. The author is indebted to Donald H. Kausler for helpful guidance and suggestions during the course of this study.

[2] Now at the Institute of Human Learning, University of California, Berkeley.

Within the factorial design in this experiment, a potential source of positive transfer, interlist response differentiation, may also be identified. If the successive PA lists have disparate levels of response M (H_1-L_2 or L_1-H_2), *S*s may find it easy to differentiate the two response systems and may therefore inhibit the List 1 interference effects present during List 2 acquisition. Successive lists of similar levels of response M (i.e., H_1-H_2 and L_1-L_2) do not have this source of positive transfer available. Interlist response differentiation has previously been discussed within the context of List 1 overlearning (Postman, 1962; Thune & Underwood, 1943) and has been invoked to explain the decline in negative transfer found as List 1 overlearning is manipulated. The positive effects attributed to response differentiation have been relatively slight (Postman, 1962) while the negative transfer effects under high-response M have been pronounced (Jung, 1963). The relative potency of the two factors leads to the prediction that the amount of negative transfer observed under the A-C paradigm should decrease, in the following order, under the H_1-H_2, H_1-L_2, L_1-L_2, and L_1-H_2 combinations of List 1 and List 2 response M.

METHOD

Design. The effects of Lists 1 and 2 response M on the magnitude of transfer were studied under two different paradigms, the experimental or A-B, A-C paradigm, and the control (for learning-how-to-learn, warm-up, etc.) or A-B, C-D paradigm. The transfer paradigms (A-C and C-D), List 1 response M (high and low), and List 2 response M (high and low) served as the three factors in a standard 2 × 2 × 2 factorial design.

Subjects.—The *S*s were 120 volunteers from general psychology classes at Saint Louis University. Each *S* was assigned randomly to one of the eight treatments by a table of random numbers, with the restriction of equal *N*s within each block of eight *S*s.

Materials.—Paired-associate lists consisting of eight two-syllable adjectives (low intralist similarity) as stimulus terms and nonsense syllables as response terms were used for all conditions. The adjectives were selected from those previously used by Kausler and Kanoti (1963) and the nonsense syllables were selected from Archer (1960). The C-D paradigm required two sets of stimuli (low interlist similarity) for Lists 1 and 2, and both the C-D and A-C paradigms required two sets of responses.

The mean association value for responses in low-M lists was 14%; the comparable value for high-M lists was 93%. Intra- and interlist response similarity was low. All *S*s assigned to the same level of List 2 response M learned the same List 2. Identical stimuli were used for List 2 across all treatment conditions.

Procedure.—Each *S* was given standard instructions regarding PA learning and the anticipation method, followed by three trials on a warm-up list consisting of three adjective nonsense-syllable pairs. Response terms on the warm-up list approximated the mean M that *S* would experience on List 1. All lists were presented on a Lafayette memory drum at a 2:2-sec. rate, with a 6-sec. intertrial interval. In order to reduce the possibility of serial learning, each list was presented in three random orders.

The experimental tasks consisted of List 1 and List 2 learning. Instructions regarding each of the tasks were not given until it was about to be initiated, nor did the instructions forewarn *S* of the succeeding parts of the experiment. List 1 learning was carried to a criterion of one perfect repetition. All *S*s were given 15 acquisition trials on List 2, with practice being initiated approximately 2 min. after the List 1 criterion had been reached.

RESULTS

List 1 learning.—While only two variables (List 1 response M and paradigms) differentiated the treatments to which *S*s were assigned, the List 1 learning data were analyzed as if all eight treatments were actually present on List 1. This was done in order to determine whether the List 2 transfer effects could be attributed to initial differences in learning abilities between the groups of *S*s. A 2 × 2 × 2 analysis of variance (trials to criterion) revealed no significant sources of variance except that for List 1 response M with an F (1,112) = 78.79, $p < .001$. No other main effects or interactions approached significance. The mean trials to criterion for low-M lists was 26.48, while the comparable value for high-M lists was 13.82.

List 2 learning and transfer effects.—List 2 performance was analyzed with respect to the number of correct anticipations on Trials 1–5 and separately for Trials 1–15. Table 1 presents the mean correct anticipations and the mean difference scores (A-C − C-D) for each of the response-M combinations. A 2 × 2 × 2 analysis of variance involving correct anticipations revealed List 2 response M to be significant with Fs (1,112) = 63.96 and 96.16, $p < .001$, for Trials 1–5 and Trials 1–15, respectively. Regardless of the level of response M on List 1, high response-M lists were learned more rapidly than low response-M lists. In addition,

TABLE 1

MEAN CORRECT RESPONSES ON LIST 2

Paradigm	Response M Combinations			
	H_1-H_2	H_1-L_2	L_1-L_2	L_1-H_2
	Trials 1–5			
A-B, A-C				
M	7.93	2.53	5.00	14.67
SD	6.36	3.23	3.81	6.46
A-B, C-D				
M	13.27	4.87	6.20	12.40
SD	4.47	4.33	3.53	5.85
Mean Difference (A-C − C-D)	−5.34	−2.34	−1.20	+2.27
	Trials 1–15			
A-B, A-C				
M	68.27	39.93	49.80	85.93
SD	18.37	19.42	17.71	14.00
A-B, C-D				
M	84.93	50.13	59.40	82.47
SD	13.68	15.95	16.48	15.49
Mean Difference (A-C − C-D)	−16.66	−10.20	−9.60	+3.46

the effect of List 2 response M was independent of the paradigm treatment as indicated by the lack of a significant interaction effect between List 2 response M and paradigms ($F < 1$ for both analyses).

The effect of List 1 response M was also significant for both analyses, F (1,112) = 6.78, $p < .05$, and 7.58, $p < .01$. However, the relationship was the inverse of that for List 2 response M. If *S* learned List 1, comprised of high-M (H_1) responses, performance on List 2 was retarded (relative to *S*s experiencing List 1 comprised of low-M (L_1) responses. This effect, although it was present under both the A-C and C-D paradigms, was almost wholly accounted for by the A-C treatments. A significant interaction effect, F (1,112) = 5.54, $p < .05$, five-trial analysis, between List 1 response M and paradigms supports the legitimacy of this conclusion. Under the C-D paradigm, means, pooled over levels of List 2 response M, were comparable irrespective of the level of List 1 response M. The same comparison under the A-C paradigm shows the H_1 means to be disproportionately smaller than the L_1 means. The relevant means for these comparisons (pooled across List 2 response M) were: C-D high, 9.07; C-D low, 9.30; A-C high, 5.23; A-C low, 9.84. The above treatment means, in the same order, for Trials 1–15 were: 67.53, 70.94, 54.10, and 67.87. The List 1 Response M × Paradigms interaction effect did not reach significance, F (1,112) = 2.76, $p < .10$, for Trials 1–15. However, as will be noted, the effect of high List 1 response M on List 2 performance was still much stronger under the A-C paradigm.

The main effect for paradigms reached statistical significance, F (1, 112) = 7.00, $p < .01$, for Trials 1–15. List 2 performance was better under the C-D paradigms, i.e., negative transfer occurred, except for the L_1-H_2 C-D treatment where A-C L_1-H_2 performance was better. The same results were obtained for Trials 1–5 except that the paradigm main effect did not reach statistical significance ($F = 3.16$, $p < .10$). None of the interaction effects involving List 2 response M or paradigms approached significance.

Finally, analyses were made to determine whether the transfer effects under the A-C paradigm were a function of response M. Mean difference scores were obtained by subtracting the mean correct anticipations (at each criterion) for the C-D condition from each A-C mean at the same level of List 1 and 2 response M. Thus, an estimate of the specific transfer effect was obtained for each A-C treatment.

Separate *t* tests were performed, as in the Jung (1963) study, comparing the four A-C mean difference scores (see Table 1) for Trials 1–5 and separately for Trials 1–15. The error mean squares of the analyses involving correct anticipations were used as estimates of error variance. For the data from Trials 1–5 mean differences in transfer effects of 2.57 and 3.38 were needed to reject the null hypothesis at the .05 and .01 levels ($s_{\bar{x}} = 1.31$, $df = 112$), respectively. The H_1-H_2 treatment was significantly different ($p < .01$) from the L_1-L_2 and L_1-H_2 treatments and the L_1-H_2 treatment also differed significantly from the H_1-L_2 and L_1-L_2 treatments ($p < .01$).

For the data from Trials 1–15 ($s_{\bar{x}} = 4.41$, $df = 112$), differences of 8.64 and 11.38 were needed to reject the null hypothesis at the .05 and .01 significance levels. For these comparisons, the L_1-H_2 treatment differed significantly ($p < .01$) from the remaining three treatments.

DISCUSSION

The present experiment strongly implies that high List 1 response M is a potent source of negative transfer in the A-C paradigm. This effect may be interpreted by Postman's "response-mediation" conception of transfer and retroaction. Postman (1961) stated that the learning of the A-C list may be mediated by the "B" List 1 responses in the manner A-(B)-C. With high interlist response similarity the mediation would be successful because of the usually strong associative strength between similar responses. However, with decreasing B-C similarity, the attempt to mediate the learning of the A-C list would become increasingly difficult because of the weak associative strengths of the B and C responses. In fact, with complete response dissimilarity, the strong associates to the B (List 1) responses would compete or interfere with the learning of A-C.

Regarding the relationship between List 1 response M and its effects on the capacity to mediate the learning of List 2, by definition, high-M responses evoke many associates. Low-M responses evoke few associates. Negative transfer in the A-C paradigm (low response similarity) should thus be a direct function of the number of associates elicited by the List 1 responses. When the List 1 response elements were of high M, List 2 learning should be retarded because *S* would have to extinguish the B responses and their associates. Less competition would be expected when List 1 was comprised of low-M responses because these responses elicit fewer associates to compete during the acquisition of List 2.

The adoption of the response-mediation hypothesis, however, demands that the interpretation of the nature of response competition be extended. Rather than being limited to competition from List 1 responses, the interference effect now must be considered to be a function of the number of associates elicited by the individual List 1 response elements. In fact, the latter source of negative transfer, in view of the pronounced difference in transfer between the high and low List 1 response-M treatments, may well be assumed to be the most potent source of response competition.

The decreasing negative transfer for the H_1-L_2 and L_1-L_2 treatment and the positive transfer obtained with the L_1-H_2 treatment presents some interesting implications for transfer theory. As predicted, the L_1-H_2 treatment had the highest level of performance on List 2 and the least negative transfer (relative to the other A-C treatments). The fact that the net transfer effect for this treatment was actually positive suggests that the differentiation factor contributed a facilitative effect considerably greater than the interference factor. The magnitude of transfer found with each A-C treatment, then, was a joint function of the relative strengths of the differentiation and interference factors.

Further discussion regarding the effects of List 1 response M is warranted. While the interpretation given fits nicely into Postman's (1961) response-mediation interpretation of transfer and retroaction, the locus of the interference effect cannot be unequivocally attributed to List List 1 response M. The experimental treatments did differ markedly in List 1 response M, but they also differed in degree of List 1 learning (in terms of trials to criterion). This fact, in principle, could account for the positive relationship found between List 1 M and negative transfer, either by greater warm-up or learning-how-to-learn in the L_1 treatments, or, by some as yet unexplained relationship between trials to criterion on List 1 and negative transfer. More complete discussions of this problem are presented by Anderson (1963) and Underwood (1964).

This interpretation is not favored for two reasons: (*a*) The positive effects of List 1 overlearning in reducing negative transfer have been small (e.g., Postman, 1962), relative to the pronounced effects obtained in the present experiment; and (*b*) The effects of List

1 practice, if attributable to warm-up or learning-how-to-learn, presumably should also have been present in the C-D paradigm treatments. In this case, the effects of List 1 practice would have been controlled by the use of difference-score measures (A-C − C-D) to assess transfer. Additionally, this discussion is not applicable to the interlist response-differentiation factor in that the effect was present both in the H_1 (H_1-L_2) and L_1 (L_1-H_2) A-C treatments.

REFERENCES

Anderson, N. H. Comparisons of different populations: Resistance to extinction and transfer. *Psychol. Rev.*, 1963, **70**, 162–174.

Archer, E. J. A re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychol. Monogr.*, 1960, **74**(10, Whole No. 497).

Jung, J. Effects of response meaningfulness (*m*) on transfer of training under two different paradigms. *J. exp. Psychol.*, 1963, **65**, 377–384.

Kausler, D. H., & Kanoti, G. A. R-S learning and negative transfer effects with a mixed list. *J. exp. Psychol.*, 1963, **65**, 201–205.

Merikle, P. M., & Battig, W. F. Transfer of training as a function of experimental paradigm and meaningfulness. *J. verbal Learn. verbal Behav.*, 1963, **2**, 485–488.

Postman, L. The present status of interference theory. In C. Cofer (Ed.), *Verbal Learning and verbal behavior*. New York: McGraw-Hill, 1961. Pp. 152–179.

Postman, L. Transfer of training as a function of experimental paradigm and degree of first-list learning. *J. verbal Learn. verbal Behav.*, 1962, **1**, 109–118.

Postman, L. Does interference theory predict too much forgetting? *J. verbal Learn. verbal Behav.*, 1963, **2**, 40–48.

Thune, L. E., & Underwood, B. J. Retroactive inhibition as a function of degree of interpolated learning. *J. exp. Psychol.*, 1943, **32**, 185–200.

Underwood, B. J. Degree of learning and the measurement of forgetting. *J. verbal Learn. verbal Behav.*, 1964, **3**, 112–129.

Underwood, B. J., & Schulz, R. W. *Meaningfulness and verbal learning*. Philadelphia: Lippincott, 1960.

(Received May 16, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 3, 270–276

OPTIMAL RESPONDING IN MULTIPLE-CUE PROBABILITY LEARNING[1]

CAMERON R. PETERSON,[2] KENNETH R. HAMMOND, AND DAVID A. SUMMERS

University of Colorado

57 Ss participated in a 3-cue, probability-learning experiment for 200 trials. The response system of Ss was compared with the optimal response system as defined by a linear multiple-regression equation. Results indicated that Ss fell only a little short of the optimal response strategy. The response-criterion correlation of about .73 corresponded to an optimal value of .83. Both (a) the response multiple R of about .90 and (b) the response-optimal response correlation of about .85 corresponded with optimal values of 1.0. Response-cue weightings of about .41, .33, and .21 corresponded with optimal values of .50, .33, and .17. Confidence intervals (95%) of the response-cue weightings remained appropriately separated after the 1st 40 trials.

Two distinguishable research approaches have been applied to the classical probability-learning experiment. The descriptive approach focuses on the performance of *S*s, often to test predictions derived from a statistical learning model. These are models of actual behavior. The normative approach compares the performance of *S*s with ideal performance, where ideal performance is specified by a rational model. Whereas the descriptive approach is concerned with actual behavior, the normative approach is concerned with the discrepancy between actual and ideal behavior (cf. Estes, 1961).

Multiple-cue, probability-learning research has, for the most part, been descriptive (e.g., Hursch, Hammond, & Hursch, 1964; Smedslund, 1955; Summers, 1962; Uhl, 1963). The present experiment, by contrast, is normative. It investigates man's capacity as an intuitive statistician in the multiple-cue, probability-learning experiment (Brunswik, 1956, pp. 80–82, 127–130).

Optimal behavior in the classical probability-learning task is to minimize expected error by predicting, on every trial, the most frequent event. There is also a strategy that will minimize expected error in the multiple-cue, probability-learning task. This strategy minimizes the square of the discrepancy between the prediction of the criterion and the actual criterion.

The optimal strategy[3] is incorporated in the multiple-regression equation,

$$\hat{Y} = a + b_1X_1 + b_2X_2 + \ldots, \quad [1]$$

where $\hat{Y}$ is the predicted value of the criterion, Y. Equation 1 minimizes the expected squared discrepancy between Y and $\hat{Y}$. The $\hat{Y}$ is comprised of a constant value, the intercept a, plus the weighted sum of the cue values when each cue, X_i, is weighted by a b coefficient, b_i. The constants, a and b, are selected such that the expected squared discrepancy between $\hat{Y}$ and Y is minimized. Thus, the criterion of minimizing expected squared error defines the optimal strategy for a multiple-probability learning task as responding with $\hat{Y}$s which are specified by Equation 1. This strategy is analogous to the maximizing strategy of a simple probability-learning task. Both strategies have the effect of minimizing expected error; the multiple-regression equation, 1, extends the principle of minimizing error from the two-choice task to the multiple-cue task.

---

[1] The research reported here was undertaken in the Behavior Research Laboratory, Institute of Behavioral Science, University of Colorado, and is Publication No. 49 of the Institute. The research was supported by a Research Grant (M-4977) from the National Institute of Mental Health. The authors are indebted to F. J. Todd for helpful suggestions.

[2] Now at the University of Michigan.

[3] The present discussion will restrict itself to the *linear* multiple probability-learning task, in which all of the systematic covariance between the cues and the criterion lies on a linear regression plane, and the remaining covariance is error.

The purpose of the present experiment was to measure several characteristics of the response system of *S*s in a multiple-cue, probability-learning experiment, and to compare these characteristics with those of an *optimal* response system as defined by the appropriate multiple-regression equation, 1, and thus to evaluate *S*'s capability as an intuitive statistician.

METHOD

Stimulus array.—The multiple-cue, probability-learning paradigm calls for several cue variables which are probabilistically related to the criterion variable. Three cues, each with a different b coefficient, were used in the present experiment. Cue and criterion values comprised the 200 multivariate trials of the experiment.

The following procedure generated cue values, preliminary criterion values, and corrected criterion values. (*a*) Two hundred numbers between 1 and 10 were selected from a random number table for each cue, yielding rectangular distributions with approximately zero correlations between cue variables. (*b*) The preliminary criterion for each trial was assigned a value equal to the weighted sum of the three cues (X_a, X_b, and X_c) according to the following parametric equation

$$Y = 0 + 1/6X_a + 1/2X_b + 1/3X_c. \quad [2]$$

This procedure yielded a criterion variable of approximately normal distribution which was linearly determined by the three cues. (*c*) The task was made probabilistic by reducing the dependence of the criterion upon the cues. One of the nine operations (+3, +2, +1, 0, 0, 0, −1, −2, −3) was applied to each criterion value, depending upon which number was found in a random number table. The criterion was then rounded off to the nearest whole number between 1 and 10. The resulting relation between the cues and the criterion was no longer deterministic; the stimulus multiple R was .83.

The stimulus array of 200 trials created by *a*, *b*, and *c* possessed approximately the same characteristics that were generated by the parametric equation, 2. The intercept was .02 and the respective b coefficients were .15, .49, and .36.

Response measures.—The three response b coefficients served as measures of cue weighting. This was a departure from the measure traditionally used—the correlation coefficient between cue and response. Whereas the traditional concern has been with the degree to which the response system provides a "mirror image" of the stimulus system (Brunswik, 1956, pp. 50 and 141), the present experiment investigated the extent to which the response system corresponded with the optimal response system.

Certain correlations do, however, provide a comparison of actual behavior with the optimal behavior of Equation 1. These correlations were used as response measures in the present experiment: (*a*) the multiple R of the response system; (*b*) the correlation between response and the parametric regression plane defined by Equation 2; and (*c*) the achievement correlation between responses and criteria. The rationale for using these measures is as follows. The optimal equation, 1, is a statistic which is an estimate of the parametric equation, 2. As the observation base (number of trials upon which the optimal equation, 1, is based) increases, its constants should stabilize and come to approximate the parametric constants of Equation 2. Thus, in the case of optimal behavior, the measures *a* and *b* should approach 1.0. As a consequence, the measure *c* should fluctuate around the stimulus multiple R of .83; the multiple R is equal to the correlation between Y and $\hat{Y}$ (cf. Hursch et al., 1964, p. 50).

Subjects.—Fifty-seven students from an introductory psychology course served as *S*s.

Apparatus.—In order to conduct group testing, a method was needed by which both cue and criterion values could be displayed to all *S*s. For this purpose, the apparatus consisted of a 3 × 4 ft. plywood board on which three vertical columns served as cues. The columns, labeled A, B, and C, were numbered from 1 to 10. Directly below these columns was a single horizontal row, also numbered from 1 to 10, which served as the criterion. The board was constructed with hooks protruding over each number so that 3-in. diameter rings could be positioned by *E*, on the board, encircling the appropriate cue and criterion values. The *S*s recorded their responses throughout the task in answer booklets.

Procedure.—The *S*s' primary task was to

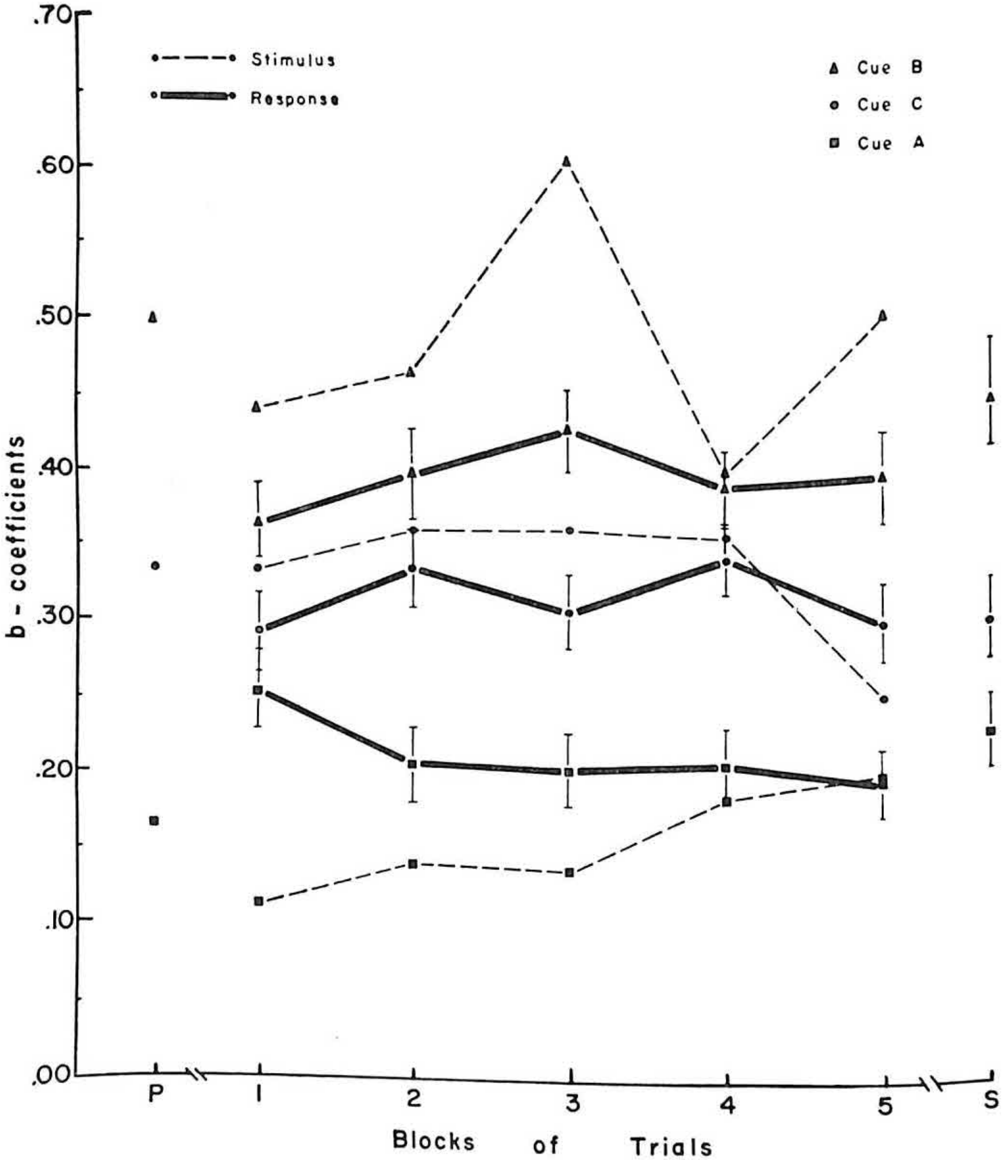


FIG. 1. Comparison of stimulus and response cue weightings as a function of trial blocks. (Vertical lines indicate the respective 95% confidence intervals. The extreme left-hand column, P, represents the parametric *b* coefficients. The extreme right-hand column, S, represents the subjective cue weightings at the end of the training blocks.)

predict, as closely as possible, the criterion value for each of 200 trials. Each trial consisted of three steps: (*a*) the three cue values were signified by rings placed over the appropriate numbers on the display board; (*b*) each *S* recorded his criterion prediction in his answer booklet; (*c*) *S*s were shown the correct criterion value by a ring placed over the appropriate number on the horizontal bar.

Prior to the prediction task, *E* read to *S*s nontechnical instructions which focused upon the following properties of the task: (*a*) the task was probabilistic, thus preventing perfect accuracy of prediction; (*b*) the task was additive, i.e., error could be minimized if each prediction was based upon the properly weighted sum of the cue values; and (*c*) the cue weightings remained stationary over the 200 trials despite occasional random fluctuations. It was pointed out to *S*s that this information paralleled that employed by an "ideal statistician" with whom their performance was to be compared.

A secondary task involved subjective judgment of cue weightings. Following each block of 40 trials, *S*s were instructed to record the *rank* of the cues (Columns A, B, and C) in determining the criterion (horizontal bar) value. The ranking was intended merely to focus *S*s upon the notion of differential cue weighting; these data were thus not analyzed. Immediately following the completion of 200 trials, *S*s were instructed to indicate their final judgment of the cue weightings. The *S*s recorded their judgments by partitioning into three areas a 100-point scale printed in the answer booklet. The relative magnitude of the respective areas was to reflect the relative weighting of the respective cues.

RESULTS

Comparison of optimal and response constants.—Block-by-block comparisons of task and response constants are presented in Fig. 1 and Table 1. Figure 1 focuses upon differential cue weighting. The left-hand column presents the parametric *b* coefficients, the cue weightings of Equation 2 used in generating the stimulus array. The actual block-by-block stimulus cue weightings, the dashed lines, fluctuate randomly around the corresponding parametric values. Optimal cue weightings would approximate cumulative values of those block-by-block values connected by the dashed lines. The solid response curves indicate that the response cue weightings quickly separated in the appropriate directions, but that the separation was of less than optimal magnitude. The response curves are essentially horizontal from Block II to the end of training. One exception is a dip in Block IV for the highest cue weighting. This dip, however, corresponds with an even sharper dip in the stimulus *b* coefficient. With this exception, the 95% confidence intervals separate after Block I and remain separated.

TABLE 1

STIMULUS AND RESPONSE INTERCEPTS

Block	I	II	III	IV	V
Stimulus Intercept	.465	.133	−.725	.343	.229
Mean Response Intercept	.418	.377	.278	.410	.655
σ_M	.010	.078	.079	.114	.100

Like the response weightings, the subjective cue weightings obtained after the training trials separated in the appropriate directions, but the magnitude of separation was less than optimal. The subjective weightings appear somewhat higher than the corresponding response weightings, perhaps because the subjective weightings were forced to sum to 1 whereas the response weightings typically summed to a little less than 1.

The other comparison of subjective and optimal constants, the intercept, is presented in Table 1. The block-by-block stimulus intercept fluctuated around the parametric value of zero; the mean response intercept remained slightly above zero; and the standard error of the mean response intercept was generally small.

Response correlations.—The corre-

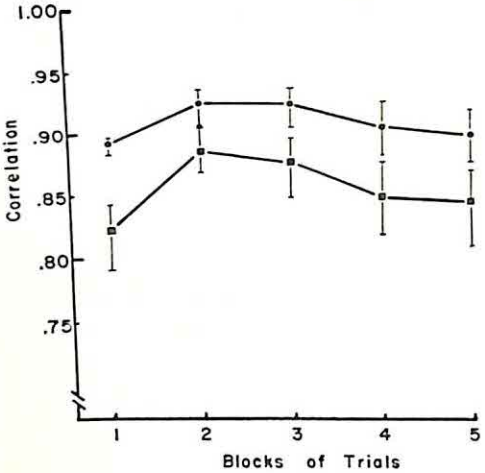


FIG. 2. Response multiple *R*s (circles) and correlations between response and parametric regression plane (squares) as a function of blocks of trials. (Vertical lines indicate the respective 95% confidence intervals.)

lations[4] in Fig. 2 and 3 permit a further comparison of *S*s' behavior with ideal behavior. The top curve of Fig. 2 represents the response multiple *R* during each block. This measure should remain at approximately 1.0 for optimal responses after attainment of stable cue weightings. The mean response multiple *R* closely approximated this; it remained near .90 throughout the blocks of trials. The lower curve of Fig. 2 represents the correlation between responses and the parametric regression plane. When the optimal equation, 1, stabilizes at the parametric equation, 2, optimal responses should lie on the parametric plane, resulting in a correlation of 1.0. The response curve falls short of 1.0; it fluctuates around .85.

The solid curve of Fig. 3 is the achievement coefficient. As optimal responses approach Equation 2, the achievement correlations should approach the dashed curve of Fig. 3. This curve represents the block-by-block correlations between the parametric regression plane and the actual criterion values; it fluctuated around the stimulus multiple *R* of .83. The achievement of *S*s generally paralleled, but fell about .10 unit below, the optimal achievement.

DISCUSSION

The purpose of the experiment was to compare the response system of *S*s with the optimal response system in a multiple-cue, probability-learning task. For each analysis the result was the same: The response system of *S*s fell a little short of the optimal response system. Although *S*s consistently deviated from

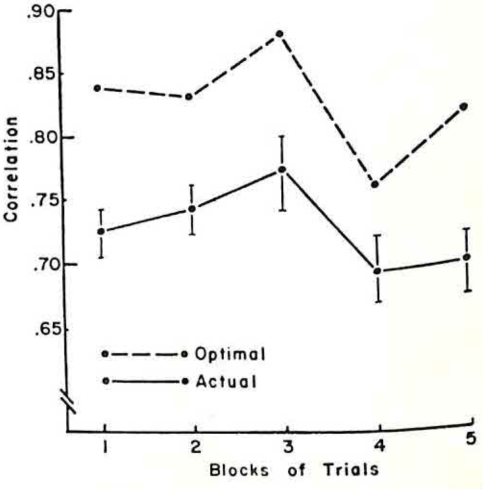


FIG. 3. Correlations between criteria and parametric regression plane (dashed line) and correlations between responses and criteria (solid line) as a function of blocks of trials. (Vertical lines indicate the respective 95% confidence intervals.)

---

[4] Response correlations of Fig. 2 and 3 have been averaged by means of the *z* transformation. This transformation is not strictly legitimate in the case of the multiple *R*, but it was the most reasonable procedure available which would yield confidence intervals comparable to those describing product-moment *r*'s. Empirical support for the use of the *z* transformation upon multiple *R*s is that untransformed median *R*s in no case differed from the reported means by as much as .02.

optimal performance, the amount of deviation was generally small. Overall achievement was high during all blocks of trials. It fluctuated only about .10 correlation point below the optimal level. The linear dependence of responses upon cues was also quite high; the average response multiple R fluctuated near .90, only .10 correlation point below the optimal. The correlation of responses with the parametric regression plane was only about .05 correlation point below the response multiple R; therefore nearly all of the linearity in the response system was based upon the cues in an appropriate manner. Additional evidence for this point resides in the response b coefficients presented in Fig. 1. The mean response b coefficients separated in the appropriate directions as early as Block I, the 95% confidence intervals generally remained appropriately separated thereafter. However, the separation was less than optimal during all blocks.

The finding that actual performance was in the optimal direction, but was of somewhat less than optimal magnitude is congruent with the results of certain other experiments which have studied *S*s as intuitive statisticians. Static subjective probabilities have been shown to underestimate high and overestimate low relative frequencies (e.g., Attneave, 1953; Preston & Baratta, 1948). Experiments using Bayes' theorem as a normative model have shown either that (*a*) the revision of subjective probabilities is in the correct direction, but of insufficient magnitude (Edwards & Phillips, 1964; Peterson, Schneider, & Miller, 1965), or that (*b*) revision of subjective probabilities is only partially sensitive to optimal revision (Peterson & Miller, 1965). The matching finding of the classical probability-learning experiment is, of course, a similar finding (Grant, Hake, & Hornseth, 1951). Matching lies between optimal and random performance.

The probability-matching model provides an interesting alternative to the normative model from which to view experiments such as the present one. A probability match between stimulus and response multidimensional scatterplots implies, as does the normative model, that the response intercept and response b coefficients match corresponding stimulus constants. However, predictions from the two models differ with respect to the response multiple R. Whereas the normative model implies a response multiple R of near 1.0, the probability-matching model predicts a match between stimulus and response multiple *R*s. According to the probability-matching model, *S* should match the random variance as well as the systematic variance in the stimulus array. Thus, the probability-matching model predicts that the response multiple R should reach (asymptotically) .83 in the present experiment. According to Fig. 2, the confidence intervals of the response multiple R were above .83 during all blocks of trials. This result means that behaving in accord with the probability-matching model does not account for the failure to reach optimal achievement in the present experiment.

Finally, responses in this experiment did not result in typical learning curves. All measures of performance apparently reached asymptotic level by Block II. This is a sharp break from the usual finding of slow learning in multiple-cue learning experiments (e.g., Smedslund, 1955). A possible explanation is that *S*s in the present experiment were instructed that the task was linear and stationary; such instruction may have resulted in faster learning by eliminating many erroneous hypotheses before the blocks of trials began.

REFERENCES

ATTNEAVE, F. Psychological probability as a function of experienced frequency. *J. exp. Psychol.*, 1953, 46, 81–86.

BRUNSWIK, E. *Perception and the representative design of psychological experiments.* Berkeley: Univer. California Press, 1956.

EDWARDS, W., & PHILLIPS, L. D. Man as transducers for probabilities in Bayesian command and control systems. In C. L. Bryan & M. W. Shelly (Eds.), *Human* *judgments and optimality.* New York: Wiley, 1964.

Estes, W. K. A descriptive approach to the dynamics of choice behavior. *Behav. Sci.*, 1961, **6**, 177–184.

Grant, D. A., Hake, H. W., & Hornseth, J. P. Acquisition and extinction of a verbal conditioned response with differing percentages of reinforcement. *J. exp. Psychol.*, 1951, **42**, 1–5.

Hursch, C. J., Hammond, K. R., & Hursch, J. L. Some methodological considerations in multiple-cue probability studies. *Psychol. Rev.*, 1964, **71**, 42–60.

Peterson, C. R., & Miller, A. J. Sensitivity of subjective probability revision. *J. exp. Psychol.*, 1965, **70**, 117–121.

Peterson, C. R., Schneider, R. J., & Miller, A. J. Sample size and the revision of subjective probabilities. *J. exp. Psychol.*, 1965, **69**, 522–527.

Preston, G., & Baratta, P. An experimental study of the auction-value of an uncertain outcome. *Amer. J. Psychol.*, 1948, **61**, 183–193.

Smedslund, J. *Multiple probability learning.* Oslo: Akademick Forlog, 1955.

Summers, S. A. The learning of responses to multiple weighted cues. *J. exp. Psychol.*, 1962, **64**, 29–34.

Uhl, C. N. Learning of interval concepts: I. Effects of differences in stimulus weights. *J. exp. Psychol.*, 1963, **66**, 264–273.

(Received May 18, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 3, 277–283

SHORT-TERM, PERCEPTUAL-RECOGNITION MEMORY FOR TACHISTOSCOPICALLY PRESENTED NONSENSE FORMS[1]

RICHARD A. STEFFY[2]

Veterans Administration Hospital, Danville, Illinois

AND CHARLES W. ERIKSEN

University of Illinois

A display consisting of triangular arrangements of 3 Vanderplas and Garvin nonsense forms was tachistoscopically presented. In the center of the triangle a 4th form (cue) occurred which was identical to 1 of the 3, and S indicated which of the 3 was identical to the cue. 2 sequences of presentation were used: cue occurred 1st followed at delays of 10–700 msec. by the 3 alternatives; alternatives occurred 1st followed by the cue. There were 2 conditions; 1 where the adapting and delay fields of the tachistoscope were dark, and the other where they were illuminated. 36 *S*s were run. For dark-adapting and delay fields impairment in recognition was obtained at delays under 100 msec. for both sequences. A marked superiority of the cue followed by alternatives sequence was found in both conditions. This result was shown to be consistent with findings of other studies on short-term memory.

Models developed to account for short-term, perceptual-memory effects portray a system in which the peripheral stimulus event is held as an accurate, fully developed representation for a duration of somewhere in the neighborhood of a quarter of a second (Averbach & Coriell, 1961; Sperling, 1960, 1963). The decay of this image is assumed to be a monotonic function of time reaching an asymptote corresponding to a more permanent type of memory or storage. Recent evaluations of these positions have suggested various reservations. Posner (1963) citing work by Pollack (1959) on information-handling strategies suggests that this perceptual storage system accounts for a limited range of perceptual events. Eriksen and Steffy (1964) were unable to demonstrate the decay function in a series of studies and found response accuracy to be quite susceptible to a variety of suppressing influences other than short-term decay.

The present studies have explored further the temporal character of perceptual storage using more complex stimulus material and a forced-choice recognition procedure.

Experiment I

Method

The experimental task required *S* to identify a cue nonsense form from a group of three alternative nonsense forms whose brief presentation terminated on half the trials 10–700 msec. before the brief presentation

---

[1] This research was supported in part by the Veterans Administration Hospital, Danville, Illinois, and in part by Research Grant M-1206, National Institutes of Health. The authors wish to acknowledge their indebtedness to Roy M. Hamlin, Research Psychologist, and Robert E. Jones, Chief of Psychology Service of the Danville Veterans Administration Hospital for their extensive support and cooperation in the conduct of these experiments. Richard H. Price, Veterans Administration psychology trainee, assisted in the collection of data.

[2] Now at the Ontario Hospital, New Toronto, Toronto, Ontario, Canada.

of the cue form. On the remaining trials the cue form terminated by these time intervals prior to presentation of the alternatives. Both the cue form and the multiple-choice alternatives were presented tachistoscopically. The presentation sequences (alternatives before or alternatives after the cue form) were varied since it was thought this might provide a sensitive test as to whether short-term perceptual memory differs in kind from long-term memory. Presenting the alternatives first requires that *S* remember three forms (only two if optimum strategy is used) to compare with the cue form when it occurs. Presenting the cue first requires memory for only one form and consequently offers less possibility of forgetting due to interference effects among the memory traces.

Two conditions of viewing were employed. Under the dark condition the adaptation field and the delay field occurring between cue (alternatives) and alternatives (cue) were dark while under the light condition both adaptation and delay fields were illuminated. On the basis of previous findings (Eriksen & Hoffman, 1963; Eriksen & Steffy, 1964; Thompson, 1964) it was anticipated that the dark condition would result in luminance summation between the two stimulus fields leading to impaired performance due to contrast reduction of the forms at delays of less than 100 msec.

The *S* was introduced to the experimental task by way of a threshold or base-level determination procedure. Here the cue and alternatives were presented simultaneously in a display in which the cue was in the center of a triangular arrangement of the alternatives. The *S* was instructed that a duplication of the center figure would be found in one of the three forms encircling it. He was to designate the matching figure as left, right, or top. The exposure time needed for *S* to achieve a recognition of 90% correct was determined and this exposure duration was then used for this *S* as the duration for both the cue and the alternative fields during the experimental trials. In the experimental trials that followed, delay intervals of 10, 50, 100, 300, 500, and 700 msec. were used to separate the cue and the alternatives. During these delay periods the display reverted to the adaptation field. The order of delay interval and sequence of presentation, alternatives-delay-cue (ADC) or cue-delay-alternatives (CDA), were counterbalanced across *S*s. Each *S* received 20 trials at each delay interval for each sequence. All 240 judgments were taken in a single experimental session. The *S* was instructed that he would see figures like those in the threshold phase but that the alternative figures and the cue would be separated in time. In the CDA sequence the cue, he was told, would be seen first and be followed a split second later by the three alternatives. He must hold that figure in mind to compare it with the alternatives. Corresponding instructions were given preceding the ADC sequence.

Apparatus and materials.—The stimulus figures were taken from the nonsense forms published by Vanderplas and Garvin (1959). The cue and alternative forms were solid black and presented against a white ground. The cue form was presented in the center of the display. In the alternative display the cue and two other forms of comparable complexity and association value were placed on the points of an imaginary equilateral triangle whose center would correspond to the spatial location of the cue form. The cue occupied 45′ of visual angle and the altitude of the imaginary triangle used to locate the position of the alternative forms constituted 3.3°.

One hundred and twenty cues and corresponding alternative cards were constructed for each of the presentation sequences (CDA and ADC). Although the same set of three alternative forms was matched with each cue form in the CDA and the ADC orders, the figures varied in position. The arrangement of alternative forms was ordered so that the probability of a correct answer at any position was $\frac{1}{3}$. The stimulus figures, all taken from the midrange (8- and 12-point complexity) of the Vanderplas and Garvin forms, were assigned to six sets of 20 pairs which were given in the same sequence to all *S*s. To control for contingencies among the delay intervals the pairing of set and interval was counterbalanced across all *S*s.

A three-field Dodge-type tachistoscope was used with presentation fields 15.5 cm. square and illuminated by 3.6 apparent footcandles (ftc.). One of the three fields was used as the adaptation and delay field and the cue form was presented in one of the remaining fields and the set of alternatives in the other. In the dark condition a faint, luminous cross was used as a fixation point. In the light condition the adaptation field was illuminated by 6.5 apparent ftc. and a black cross constituted the fixation point. Under this condition the stimulus fields containing the cue and the alternatives were superimposed upon the adaptation-field luminance and the interval between stimulations constituted a return to the adapting-field luminance.

Subjects.—Twenty-four male neuropsychiatric patients at the Danville Veterans Administration Hospital were selected. They were evenly assigned to the light and dark experimental conditions on a random basis. In selecting *S*s, patients with visual defects, known neurological impairment, and poor judgment were excluded. The mean age of the sample was 38.1 yr. with an *SD* of 6.49 yr.

Results

Before examining the influence of delays, cue-alternative sequences, and illumination of delay field, several between-*S* variables were examined. A comparison was made of the exposure durations necessary for *S*s in the light and dark conditions to reach the 90% criterion of recognition accuracy on the base-level series where the cue and alternatives were presented simultaneously. It had been previously found that a tachistoscopic method that employs a dark adaptation field preceding and following the stimulus exposure, requires shorter exposure durations for the same accuracy criterion than a method employing a light pre- and postadaptation field (Eriksen & Hoffman, 1963; Eriksen & Steffy, 1964). Consequently it was expected that the exposure durations in the base-line series would be shorter for *S*s in the dark condition. The mean exposure duration in milliseconds for *S*s in the dark condition was 174.6 and 220.8 for the light condition. While the difference is in the expected direction it falls just short of significance at the 5% level for a one-tailed test, *t* (22) = 1.69.

In the experimental series half the *S*s in each condition were given the sequence ADC first. The number of correct recognitions was summed by *S*s across conditions and delay intervals for the two orders. The two order means were nearly identical, *t* (22) = .97.

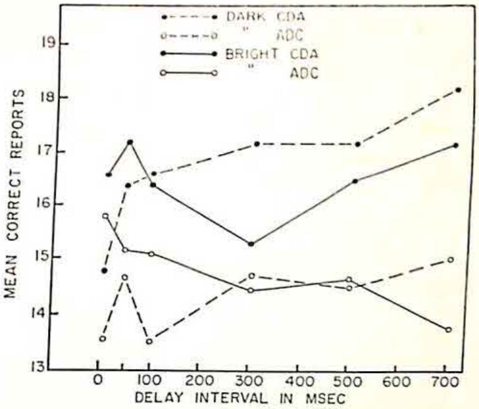


FIG. 1. Report accuracy as a function of delay time, illumination of delay field, and sequence of cue-alternative presentation in Study I.

The effect of the experimental treatments upon recognition accuracy was evaluated by a modified four-way analysis of variance: *S*s, conditions of the delay field (light or dark), delay intervals, and cue-alternative sequence. Three significant effects were obtained. Besides significant differences due to *S*s, F (22, 110) $= 5.4$, $p < .001$, the main effect for sequence is significant, F (1, 22) $= 23.2$, $p < .001$. There is also a significant Delay × Condition interaction, F (5, 110) $= 3.4$, $p < .01$.

In Fig. 1 recognition accuracy is shown as a function of delay interval for the two sequences in each of the two experimental conditions. For both the light and dark conditions the sequence, CDA, gives better recognition accuracy at all delay intervals. While there was no significant effect for the delay interval as a main effect in the analysis of variance, nonetheless introducing even a 10-msec. delay between the alternatives and the cue form results in an appreciable reduction in recognition accuracy from what is obtained when the cue and the alternatives are presented concurrently. This holds true for both conditions and for both sequences. The biggest loss occurs for the dark condition and within that condition for the ADC sequence.

The significant Delay × Condition interaction noted in the analysis of variance is seen to be quite complex. Under the dark condition the CDA sequence shows maximum loss in recognition accuracy with a delay of 10 msec. with the loss becoming progressively less until a 100-msec. delay is reached. There is also a suggestion of an improvement in accuracy from 500 to 700 msec. delay. The curve for the ADC sequence in this condition is not as clear-cut although there appears to be a tendency to follow a similar pattern of poorer recognition at the shorter delays that is partially recovered by 300 msec. In the light condition the ADC sequence suggests a progressive loss in recognition accuracy as the cue is delayed longer. However, for the CDA sequence little systematic change in recognition accuracy as a function of delay is noted. Recognition accuracy is approximately the same at 700-msec. delay as at 10- or 50-msec. delay. The only marked departure for this condition and sequence is a rather peculiar reduction in accuracy at the 300-msec. delay.

EXPERIMENT II

This experiment was an exact replication of the light condition of the preceding experiment with the exception that it was carried out upon a sample of normal *S*s. Only the light condition was replicated since the results obtained for the dark condition were commensurate with our previous findings on temporal-luminance summation and it was our purpose to get more information on the memory and decay process independent of the luminance summation effect.

Method

Procedure.—The procedure was identical to that employed for the light condition in the previous experiment.

Subjects.—Twelve male employees of the Physical Medicine and Rehabilitation Service at the Danville Veterans Administration Hospital volunteered to serve as *S*s. All were paid a fixed rate for their services. Half the group were college graduates and three others were college students working in a training capacity at the hospital. The average age of the sample was 34.25 yr. with an *SD* of 9.75.

Results

Recognition accuracy was evaluated by a three-way analysis of variance (*S*s, delays, and sequences). A t test for the order in which the CDA and ADC sequences had been administered to *S*s gave a t value of less than 1. As consistent with the previous experiment significant main effects attributable to *S*s, F (11, 55) = 5.5, $p < .001$, and to the presentation sequence, F (1, 55) = 37.3, $p < .001$, were found.

To provide a more sensitive test of delay and sequence effects as well as to determine differences between the patient and normal samples, the data in this experiment were combined with the light-condition data for the patient population in Exp. I and a modified four-way analysis of variance (patient vs. normal group, *S*s, delays, and sequences) was carried out.

Sequence and *S* effects were of course significant ($p < .001$) but there was no significant main effect attributable to the differences between the patient and normal group, F (1, 22) < 1. There is however, a significant third-order interaction between groups, delays, and sequences

($p < .001$) and a second-order interaction between delays and sequences is significant when evaluated against its error term ($p < .01$).

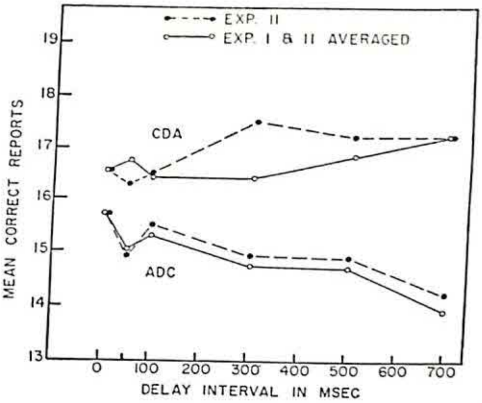


FIG. 2. Report accuracy as a function of delay time and sequence of cue-alternative presentation for Study II and for the bright field condition of Study I averaged with Study II.

The nature of these effects is seen in Fig. 2. Here recognition accuracy as a function of delay interval and sequence has been plotted for the normal group in Exp. II and also the averaged data for the normal and patient samples are plotted. If the data from the second experiment are examined first, it is noted that there is again a marked superiority in recognition for the CDA sequence at all delay intervals. The curve for the ADC sequence is quite similar to that obtained under the comparable light condition in Exp. I. The CDA sequence also shows general agreement with the light condition of Exp. I with the exception of the 300-msec. delay. In Exp. II there is an improvement in recognition accuracy at this delay interval whereas under the comparable light condition in the preceding experiment the patient *S*s showed a rather pronounced drop. This difference in performance at 300 msec. on the CDA sequences between these two samples would seem to account for the third-order interaction that was noted in the analysis of variance.

When the data for the patient sample in the light condition of the preceding experiment were combined with the data from the normal *S*s in the present study, the interaction between delays and sequences noted in the analysis of variance is seen to consist of an increasing difference in recognition accuracy between the two sequences as the delay interval is lengthened. For the CDA sequence, recognition accuracy shows a tendency to improve out to the maximum time interval investigated after an initial drop at short delays. On the other hand the ADC sequence shows a rather rapid decay in accuracy for delay intervals under 100 msec. and then a more gradual decay as the delay is lengthened.

DISCUSSION

The findings in Exp. I for the dark condition are consistent with previous data reported by Eriksen and Hoffman (1963) and Eriksen and Steffy (1964). It had been found that a stimulation sequence consisting of dark-adaptation field—luminated stimulus field—dark-adaptation field—luminated stimulus field—and dark-adaptation field, resulted in impaired recognition performance when the stimulus fields were separated by delay intervals of less than 100 msec. This impaired recognition was attributed to a reduction in the figure-ground contrast for the stimuli due to a temporal-luminance summation between the successive luminated fields. It was previously shown how a luminance summation explanation was consistent with other data in the field of vision and not only accounted for the reduced recognition with sequential stimulation at these delay intervals but also explained why durations of exposure for a given level of recognition accuracy are generally lower when a dark pre- and postadaptation is employed with single-stimulus field exposures.

In Exp. I the function obtained for the CDA sequence in the dark condition fits the luminance summation explanation very well. Recognition is poorest at the 10-msec. delay, improves at 50 and apparently approaches an asymptote by a 100-msec. delay. The function obtained for the ADC sequence is not quite as clear. The reversal between the 50- and 100-msec. delay intervals is inconsistent with this explanation but may perhaps be attributable to chance fluctuation since the rest of the function is commensurate with the explanation.

The results from neither of the conditions in Exp. I nor from the replication data of Exp. II provide support for the view that brief perceptual memory differs from long-term memory phenomena. The differences in recognition accuracy obtained between the CDA and the ADC sequences are consistent with the brief memory functions reported by Murdock (1961) and by Melton (1963). The results are not consistent with what would be expected if there was a perceptual memory comparable to a complete image of the stimulus that persisted for a brief period of time following termination of the stimulus (Sperling, 1963).

From this latter theory it would be anticipated that recognition accuracy would be equally good when the display containing the three alternative forms is presented first followed by the cue form as the reverse sequence. Under the ADC sequence a complete but rapidly decaying representation of the three alternative forms would be present at the time the cue form appears clear and definite in perception. The task would then be to match the clear cue form with the three alternatives each of which has decayed approximately the same amount. Under the CDA sequence the task would be to match a rapidly decaying perception of the cue form with three clear representations of the alternatives. The matching problem, whether one is dealing with a partially decayed set of alternatives or a partially decayed cue, would seem to be about equally difficult.

Contrasted with this view of the nature of the memory trace are the more conventional interference theories of memory. From the interference viewpoint it would be predicted that the ADC sequence would be more difficult and lead to poorer recognition than the CDA. Under the ADC sequence *S* is faced with the task of remembering three forms with the attendant interference until the cue form occurs. This represents considerably more interference than would be represented by the CDA sequence where only one form has to be remembered over the brief time span.

Further evidence that the memory phenomena obtained at these very brief intervals is comparable to memory phenomena over longer periods is observed in the very marked similarity of the functions obtained for the ADC and CDA sequences to the data reported by Murdock (1961) for words and Melton (1963) for consonants for memory intervals of from 3 to 32 sec. Murdock (1961) found that one word was remembered over an 18-sec. interval with little or no evidence of forgetting, whereas three words showed a very marked decay within the first few seconds. Similar findings were obtained by Melton (1963) for consonants. Our results show that one nonsense form is remembered with little or no impairment over the interval of 700 msec. whereas three nonsense forms show a very pronounced decay within the first 100 msec. and a tendency to decay more slowly throughout the remainder of the investigated delay intervals.

The memory intervals involved in the present experiment are very much shorter than those used by these other *E*s. Neither Murdock (1961) nor Melton (1963) found much decay at a 1-sec. interval, but both employed meaningful material and their stimuli were presented for a longer period of time. The longer duration of the stimulus may have allowed for a more firmly established trace in the Murdock (1961) and Melton (1963) studies and also the memory required may have been less complex for consonants and words than for nonsense forms.

The finding that recognition accuracy shows a drop for short delay intervals on the CDA sequence and then a gradual improvement up to 700 msec. may be attributable to retroactive interference effects that have been previously found (Eriksen & Steffy, 1964).

REFERENCES

AVERBACH, E., & CORIELL, A. S. Short-term memory in vision. *Bell Sys. tech. J.*, 1961, **40**, 309–328.

ERIKSEN, C. W., & HOFFMAN, M. Form recognition at brief durations as a function of adapting field and interval between stimulations. *J. exp. Psychol.*, 1963, **66**, 485–499.

ERIKSEN, C. W., & STEFFY, R. A. Short-term memory and retroactive interference in visual perception. *J. exp. Psychol.*, 1964, **68**, 423–434.

MELTON, A. W. Implications of short term memory for a general theory of memory. *J. verbal Learn. verbal Behav.*, 1963, **2**, 1–21.

MURDOCK, B. B., JR. The retention of individual items. *J. exp. Psychol.*, 1961, **62**, 618–625.

POLLACK, I. Message uncertainty and message reception. *J. Acoust. Soc. Amer.*, 1959, **31**, 1500–1508.

POSNER, M. I. Immediate memory in sequential tasks. *Psychol. Bull.*, 1963, **60**, 333–349.

SPERLING, G. The information available in brief visual presentations. *Psychol. Monogr.*, 1960, **74**(11, Whole No. 498).

SPERLING, G. A. A model for visual memory tasks. *Hum. Factors*, 1963, **5**, 19–31.

THOMPSON, J. H. What happens to the stimulus in backward masking? Unpublished doctoral dissertation, University of Illinois, 1964.

VANDERPLAS, J. M., & GARVIN, E. H. The association value of random shapes. *J. exp. Psychol.*, 1959, **57**, 147–154.

(Received May 19, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 3, 284–288

CHOICE REACTION WITH VARIABLE S-R MAPPING[1]

L. H. SHAFFER

Medical Research Council, Applied Psychology Research Unit, Cambridge, England

An experiment examined 2-choice reaction under fixed or variable mapping rules. Each display contained a signal, M, and a symbol representing a rule, I, that transformed it. In some conditions I and M came on in sequence. RT was always shorter under fixed than under variable I. When I was variable, presenting I or M in advance led to shorter RT, the reduction being greater for I. There were independent transition effects for I and M from one trial to the next, being greater for I than for M: RT was shorter with repetition than with alternation under variable I and conversely under fixed I. RT was longer with contra- than with homolateral mapping under fixed I or if I was variable and given in advance, otherwise it was invariant with mapping.

This is a study of the effect in performing a sequence of computations of randomly varying the class of computation within the sequence. The general program at the present time involves computation at a level no more complicated than arithmetic sums and geometric or verbal transformations. Jersild (1927), using two classes of computation—word opposites and numerical subtraction—found that an alternation of these in lists led to faster solution times than lists involving only one class: but alternations of more similar classes—add 3 or subtract 3; opposites of adjectives or opposites of verbs—led to slower solution times. The present study examines in some detail the case of similar classes, appearing in random sequence rather than in simple alternation.

Usually in choice-reaction tasks a one-one mapping of a set of signals, M, into a set of responses, R, is defined by a rule, $i - i(M) = R$. For many of these tasks R can be regarded as a systematic transformation of M by the operator i. It is a straightforward extension to introduce a set of rules, I, each of which confers a one-one mapping of M into $R - I(M) = R$. In a sequence of trials I and M can vary independently and RT can be examined as a function of each.

The present experiment was confined to the case of set sizes $I = M = 2$. Of particular interest were the transition effects when neither, one, or both of these changed from one trial to the next. The experiment was in two parts: in the first I and M appeared simultaneously in each display with no alerting signal; in the second they came on in sequence, the second component appearing with a latency $\frac{1}{3}$ sec. In a control condition a neutral alerting signal appeared at this advance interval.

METHOD

Apparatus.—The display contained two lights (M), horizontally arranged with a 6-cm. separation, and midway between them an illuminated cross of which the vertical and horizontal limbs (I) could be lit independently. The response panel contained two buttons horizontally arranged with 6-cm. separation. At each trial the display parameters were preset by E and RT was recorded by a Dekatron timer to .01 sec. An electronic device could introduce a variable delay between onset of the two display components and when used RT was measured from onset of the second component.

---

[1] Acknowledgment is made to the Medical Research Council's Naval Personnel Research Committee for research facilities.

Design and procedure.—Six groups of *S*s were used and each was tested under only one condition.

Group *2I2M* (12 *S*s): one of two lights, *M* (left or right) and one of two lines, *I* (horizontal or vertical) came on simultaneously at each trial. The horizontal line, H, defined a homolateral mapping of *M* into *R* and the vertical line, V, a contralateral mapping. In a session of 81 trials each of the four possible displays occurred 20 times in a random sequence and each of the 16 possible first-order transitions occurred five times. There was a delay of about 10 sec. between trials and no alerting signal was used.

Group 2H (6 *S*s): only the two lights varied over trials. A homolateral mapping was defined in the instruction and the horizontal line always appeared in the display. In a session of 81 trials each light occurred 40 times in random sequence and each of four possible first-order transitions occurred 20 times.

Group 2V (6 *S*s): as for 2H except that a contralateral mapping was defined and the vertical line always appeared in the display.

Group $2M \rightarrow 2I$ (12 *S*s): as for *2I2M* except that the onset of *M* always preceded the onset of *I* by $\frac{1}{3}$ sec.

Group $2I \rightarrow 2M$ (12 *S*s): the onset of *I* always preceded the onset of *M* by $\frac{1}{3}$ sec.

Group $\cdot \rightarrow 2I2M$ (12 *S*s): a dot (the intersection of H and V) always preceded the onset of *I* and *M* by $\frac{1}{3}$ sec.

The *S*s in each group received nine sessions, the first a practice session of about 10 min. and the other eight test sessions.

Instruction.—In the practice session the rules and their operation were explained. It was pointed out that rotating the plane about a vertical axis midway between the lights reversed their positions and about a horizontal axis passing through the lights left them unchanged. The *S*s were instructed to use their index fingers to press the buttons and to respond as quickly as possible without making errors.

Subjects.—The 60 *S*s were males in the age range 18–30.

RESULTS

If *I* was variable then RTs for H and V rules were separated, also RTs for the four types of transition—neither, *I*, or *M*, both changing—were separated. If *I* did not vary then RTs for repeat and alternate transitions of *M* were separated. These results averaged across sessions and across *S*s within a group are shown in Table 1. It should be noted that RT to the first trial in a session was not classified and hence discarded.

TABLE 1

MEAN RTS FOR DIFFERENT RULES, *I*, AND DIFFERENT TRANSITIONS OF *I* OR *M*

Cond.	Mean RT		Mean RT in Each Transition Category
2H	.387 sec.		$M\,\dfrac{\text{s}\ .393}{\text{d}\ .381}$ sec.
2V	.441		$\dfrac{.448}{.434}$
2I2M	H = .706	V = .698	$M\,\begin{array}{lcc} & \text{s}\quad I & \text{d} \\ \text{s} & .674 & .712 \\ \hline \text{d} & .694 & .729 \end{array}$
$\cdot \rightarrow 2I2M$	.648	.659	$\dfrac{.630 \quad .663}{.650 \quad .671}$
$2M \rightarrow 2I$	.575	.582	$\dfrac{.550 \quad .586}{.573 \quad .606}$
$2I \rightarrow 2M$	.497	.542	$\dfrac{.500 \quad .526}{.517 \quad .536}$

Note.—*I* or *M* could be the same, s, or different, d, in the transition from one trial to the next.

The results of Groups 2H and 2V were compared in an analysis of variance distinguishing groups, transitional state of *M* (same, s—different, d), sessions, and *S*s. The difference between groups was significant; $F\ (1, 10) = 5.77$, $p < .05$. The transition effect was significant; $F\ (1, 10) = 19.89$, $p < .01$. The sessions effect was significant; $F\ (7, 70) = 5.86$, $p < .001$, representing a monotonic improvement with practice. There were no significant interactions.

For Group *2I2M* a *t* test of the RT differences between H and V rules was nonsignificant; $t\ (11) = 1.17$, $p > .05$. An analysis of variance was carried out distinguishing transitional state of *I*, transitional state of *M*, sessions, and *S*s. The effects of both types of transition were significant; for I: $F\ (1, 11) = 26.44$, $p < .001$;

for M: $F\,(1, 11) = 39.56, p < .001$. The sessions effect was significant, $F\,(7, 77) = 31.29, p < .001$, representing a monotonic improvement with practice. There were no significant interactions.

A t test of the RT differences between H and V rules was carried out for each of the groups $2M \rightarrow 2I$, $2I \rightarrow 2M$, $\cdot \rightarrow 2I2M$. It was significant for the second group only, $t\,(11) = .38, p > .05$; $t\,(11) = 8.27, p < .001$; $t\,(11) = 1.17, p > .05$, for the respective groups. An analysis of variance was carried out distinguishing groups, transitional state of I, transitional state of M, sessions, and $S|$groups. The differences between groups were significant; $F\,(2, 33) = 12.23, p < .001$. The effects of both types of transition were significant, for I: $F\,(1,33) = 82.91, p < .001$; for M: $F\,(1,33) = 87.16, p < .001$. The sessions effect was significant, $F\,(7,231) = 47.12, p < .001$ representing a monotonic improvement with practice. Three interactions were significant: $I \times$ Sessions, $F\,(7, 231) = 3.63, p < .01$, $M \times$ Sessions, $F\,(7, 231) = 2.09, p < .05$, and I × M × Sessions, $F\,(7,231) = 2.65, p < .05$. None of these represent monotonic growth or diminution of the transition effects over sessions.

A t test was carried out on the RT difference between Groups $2I2M$ and $\cdot \rightarrow 2I2M$. This difference was significant; $t\,(22) = 2.12, p < .05$.

The error scores represented less than 1% of all responses and so no group comparisons were attempted. However an interesting result is contained in the classification of errors into transition categories, as shown in Table 2. The transition effect on RT is clearly mirrored by the errors made in the transition categories.

TABLE 2

TOTAL ERRORS FOR DIFFERENT RULES, I, AND DIFFERENT TRANSITIONS OF I OR M

Cond.	N	Total Errors		Errors in Each Transition Category		
2H	6	28		M s	13	
				M d	15	
2V	6	27			11	
					16	
					I s	I d
$2I2M$	12	H=44	V=22	M s	3	16
				M d	8	39
$\cdot\rightarrow 2I2M$	12	28	30		4	9
					14	31
$2M\rightarrow 2I$	12	61	52		16	24
					22	51
$2I\rightarrow 2M$	12	17	23		2	15
					3	20

Note.—s = same; d = different.

DISCUSSION

If class of computation is equated with mapping rule then the results show that RT is longer when the class is variable than when it is fixed over a sequence. Jersild's result is thus confirmed for the case when the classes are similar. It remains, of course, to define the notion of similarity and explore the generality of the finding.

In this paper the notion of computation is only formally specified as a transformation of a datum. It includes grammatical transformations, arithmetic operations and, in extension, the application of an algorithm to a problem. No attempt is made here to conjecture an underlying process, instead the experiments aim at defining some of its characteristics.

Analytic distinctions can be made to separate four component of the process of computation: identify the operator, identify the datum, carry out the transformation, and produce the correct output symbol (i.e., overt response). Consider the conditions $\cdot \rightarrow 2I2M$, $2M \rightarrow 2I$, and $2I \rightarrow 2M$. Giving one component of the display in advance of the other did not reduce the size of the possible response set, hence the time saved in the response must be attributed either to identifying that component or to a partition of the set of possible transformations $I(M)$. If the transition effect were a function of identifying a display component or of a partition, then the effect should have been removed or diminished for the component given in advance. Since the pattern of transition effects was in fact invariant over all groups it was a function of the transformation. Finally note that response, R was unchanged in a transition if both I and M were unchanged or if both changed, conversely R changed if I or M but not both changed. The two categories of transition in which R was invariant contained the shortest and the longest RTs. Thus the transition effect was not a function of the retrieval of an overt response, hence transformation and retrieval are distinct.

The parameters of the transformation $I(M)$ seem to have had independent transition effects and the effect of changing I was always larger than the effect of changing M—even when I was presented in advance. What is more the relative magnitude of these transition effects corresponded to the relative reductions in RT obtained by presenting I or M in advance.

Comparing results for the conditions when I was fixed and when it was variable, three points may be noted. The first is that the transition effects were opposite in the two cases. Repetition led to a reduction in RT when I was variable and to an increase in RT when it was fixed. The latter represents an effect analogous to the "gamblers fallacy" observed in choice guessing (Jarvik, 1951). Bertelson (1961) showed that with self-paced responding in two-choice reaction under fixed mapping repetition was faster than alternation and that the effect was reduced by interposing a $\frac{1}{2}$-sec. interval between trials. Since in all the present experiments there was a 10-sec. interval between trials it seems that the transition effect favoring repetition is a joint function of intertrial interval and any factor affecting difficulty of choice.

The second point is that with fixed I, RT was shorter under homolateral (H) than under contralateral (V) mapping. It is well established that compatibility factors, including the geometric relationship between stimuli and responses, affect RT (Fitts & Seeger, 1953). However, with variable I the difference in RT between H and V trials disappeared under all conditions except $2I \rightarrow 2M$. In the latter case, in which I was given in advance, the RT difference between H and V trials was almost as large as between Cond. 2H and 2V. It is as though when I was variable S selected at each trial from a set of equally difficult transformations: when I was fixed or known in advance the null transformation could be considered as a special class and was easier to compute. Note again that Bertelson (1963) found a larger transition effect for low-compatible than for high-compatible mapping.

The third point is that RT with fixed I was always shorter than with variable I even when I was given in advance. This difference together with the ones already mentioned suggests that, in the terminology of computers, S uses a different type of program for response selection in the two cases.

The difference in RT between Cond. $2I2M$ and $\cdot \rightarrow 2I2M$ indicates the effect of time uncertainty of the stimulus (Bertelson, 1961; Klemmer, 1956). The result throws no new light on a familiar phenomenon which suggests that the program for response selection, or some part of it, is not uniformly available in time and must be mobilized in each trial.

The pattern of errors in the transition categories can be regarded as an attempt by S to maintain a fixed rate of responding as the ease of transition changed. In this case it is to be expected that an errorless performance would show larger transition effects in RT.

REFERENCES

Bertelson, P. Sequential redundancy and speed in a serial two-choice responding task. *Quart. J. exp. Psychol.*, 1961, **13**, 90–102.

BERTELSON, P. S-R relationships and reaction times to new versus repeated signals in a serial task. *J. exp. Psychol.*, 1963, **65**, 478–484.

BERTELSON, P., & BOONS, J.-P. Time uncertainty and choice reaction time. *Nature*, 1960, **187**, 531–532.

FITTS, P. M., & SEEGER, C. M. Stimulus-response compatibility: Spatial characteristics of stimulus response codes. *J. exp. Psychol.*, 1953, **46**, 199–210.

JARVIK, M. E. Probability learning and a negative recency effect in the serial anticipation of alternative symbols. *J. exp. Psychol.*, 1951, **41**, 291–297.

JERSILD, A. T. Mental set and shift. *Arch. Psychol., N. Y.*, 1927, **89**, 1–81.

KLEMMER, E. T. Time uncertainty in simple reaction time. *J. exp. Psychol.*, 1956, **51**, 179–184.

(Received May 11, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 3, 289–293

SIZE CUES AND THE ADJACENCY PRINCIPLE[1]

WALTER C. GOGEL[2]

Civil Aeromedical Research Institute, Oklahoma City, Oklahoma

The adjacency principle asserts that the dominant cues in the determination of the perceived position of an object in a configuration of objects are the distance cues which occur between it and adjacent, not displaced, objects. The purpose of the present study was to apply this principle to the perception of relative depth from size cues. For this purpose, a visual situation was produced in which the perceived depth position of a playing card would differ depending upon the relative effectiveness of the size cue between this card and other cards. In agreement with the adjacency principle it was found that the size cue between adjacent cards was more effective than the size cue between displaced cards in the determination of perceived relative depth.

A recent report has identified a factor which has been demonstrated to be effective in the perceptual organization of the three-dimensional world (Gogel, 1963b). This factor labeled the "adjacency principle" asserts that the dominant cues in the determination of the perceived size or position of an object in a configuration of objects are the size or distance cues which occur between it and adjacent, not displaced, objects. The adjacency principle has been demonstrated mainly with respect to the cue of binocular disparity. The present study will apply the adjacency principle to the perceived depth resulting from size cues. It is predicted from the adjacency principle that size cues to relative depth between adjacent rather than displaced objects are the size cues most effective in determining apparent depth positions.

A convenient method of testing the application of the adjacency principle to a particular cue system involves modifying the perception from that which would normally occur with this cue system. The logic of this method and its relation to the present study are illustrated in Fig. 1.

Figure 1a is a top-view drawing of the physical positions of three objects (A, B, and D), all at the same distance from O. The dotted lines represent lines of sight (directions from O to the objects). Object A is directionally closer to Object B than to Object D, i.e., $\phi_2 > \phi_1$. The adjacency principle states that the apparent depth position of A (with respect to B and D) is determined mainly by whatever depth cues occur between A and B, not between A and D.

In Fig. 1, Objects A, B, and D are similar objects with the visual angular size of A equal to that of D and with both A and D half the visual angular size of B. Therefore, according to the relative-size cue, A and D should appear at the same distance as each other with both appearing more distant than B. Thus, if all distance cues other than relative size were absent, the apparent depth positions of A, B, and D would be A′, B′, and D′ as shown in Fig. 1b. Under the circumstances indicated by Fig. 1b, it cannot be determined whether the equality of visual angular size be-

---

[1] The author wishes to thank Henry W. Mertens, Frank L. Agee, Jr., and James C. Hester for their assistance in collecting and analyzing the data of this study.

[2] Now at the University of California, Santa Barbara.

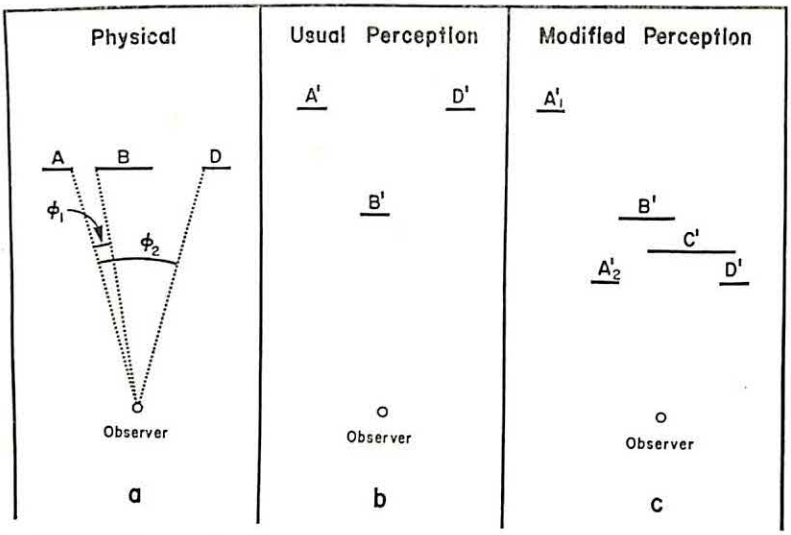


FIG. 1. Schematic top-view drawing illustrating a method of testing the application of the "adjacency principle" to size cues to relative depth.

tween A and D or the angular size difference between A and B was more significant in determining the apparent depth position of A. Therefore, the adjacency principle cannot be investigated by using the conditions illustrated by Fig. 1b.

In Fig. 1c, an additional object (Object C) has been added which, as a consequence of introducing the cue of interposition, causes B to appear more distant than D in spite of the difference in visual angle between B and D. Under these conditions, the adjacency principle can be investigated by measuring the apparent distance position of A. If A appears at A'_1 in Fig. 1c, the size cue to relative depth between A and B is dominant in determining the apparent distance position of A. However, if A appears at A'_2 in Fig. 1c, the size cue to relative depth between A and D is dominant in determining the apparent distance position of A. From the adjacency principle, it would be predicted that the apparent position of A would be more nearly A'_1 than A'_2.

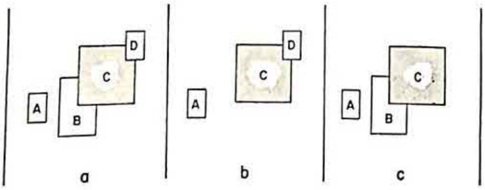


FIG. 2. Schematic front-view drawing illustrating the positions of the normal-sized playing cards (A and D), the double-sized playing card (B), and the gray square (C) for the condition in which all three cards were present (Fig. 1a) or in which one card was absent (Fig. 1b or Fig. 1c).

METHOD

Apparatus

Objects A, B, and D in Fig. 1 were similar objects, i.e., objects which differed only in retinal size, in order to produce the size cue to relative depth. In the present study, the size cue was produced by using playing cards for Objects A, B, and D. To determine whether Playing Card A appeared at A'_1 or A'_2 (Fig. 1c) it was necessary to measure the apparent depth position of Card A with respect to the other playing cards. Also, the effect of each of the other playing cards upon the apparent position of Card A was further investigated by determining the change in the apparent position of A when each of the other cards was present and absent. To accomplish this, the three displays shown in Fig. 2 were each presented one at a time to each *O*. The displays were produced by two light boxes with each box containing fluorescent lights with a diffusing glass as a front surface. Positive transparencies of a normal-sized (5.7 × 8.9 cm.) seven of spades playing card (used in producing Cards A and D) or a double-sized (11.4 × 17.8 cm.) seven of spades playing card (used in producing Card B) were fastened to the front surfaces of the light boxes. When viewed in the otherwise dark environment, the positive transparencies appeared to *O* as seven of spades playing cards. Object C was formed from a 17.8-cm. square of neutral density film and fastened on the front surface of the light boxes. Object C appeared to *O* to be a gray square partially covering Card B and partially covered by Card A and, therefore, apparently interposed in depth between Cards B and D. This was accomplished by cutting away 4% of the upper right area of Object C and 22.4% of the upper right area of Object B and carefully fitting the objects together. All of the objects (A, B, C, and D) were at the same distance (305 cm.) from *O*. The portions of the front surfaces of the light boxes not used in producing the playing cards or gray square were completely covered by black opaque cardboard. A partially reflecting mirror permitted *E* to present either of the two light boxes (one at a time) so that each display appeared directly in front of *O*. The position and orientation of the light boxes were such that, regardless of which light box was turned on, the size, brightness, direction, and distance of a particular card was unchanged. The displays presented by means of one light box are illustrated in Fig. 2a and 2b. The display shown in Fig. 2b was produced from that of Fig. 2a by covering Card B. The display presented by means of the other light box is illustrated in Fig. 2c. The luminance of Object C and of the white portions of the playing cards was .5 and 3.0 footlamberts (ftl.), respectively, as measured from the position of *O*'s eye.

The viewing position of *O* was enclosed by black cloth and contained a head and chin rest. The displays were viewed monocularly by *O* through a neutral density filter located in an eyepiece in front of his right eye. A light-adaptation surface (3 ftl.) located to the left of *O*'s position, which could be turned on or off by *E*, further insured that no objects or parts of objects except A, B, C, or D were visible to *O*.

A device for measuring the perceived depth between the several playing cards was located at the position of *O* and consisted of two vertical metallic rods (invisible to *O*) which *O* could grasp, one in each hand. The right rod was movable laterally, with the left rod fixed in position. The lateral distance between the inner edges of the rods was adjusted kinesthetically by *O* to be equal perceptually to the perceived *depth* between the designated playing cards. A centimeter rule attached to the apparatus, indicated to *E* the adjusted separation between the rods.

Procedure

The three displays (Fig. 2) were presented to each of 20 *O*s (17 men and 3 women). Each *O* had a visual acuity in his right eye (corrected if necessary) of at least 20/20. The perceived depth between playing cards which *O* was asked to indicate by means of the hand adjustments (task) and the conditions under which each task was completed were as follows:

Task 1. To indicate the perceived depth between A and D with B present (Fig. 2a).

Task 2. To indicate the perceived depth between A and D with B absent (Fig. 2b).

Task 3. To indicate the perceived depth between A and B with D absent (Fig. 2c).

Task 4. To indicate the perceived depth between A and B with D present (Fig. 2a).

Task 5. To indicate the perceived depth between B and D (Fig. 2a). The order in which these five tasks were completed was systematically varied between *O*s, with each perceived depth being measured three times.

If the size cue between Cards A and B contributed to the perception of the depth position of Card A, the perceived depth position of Card A, as judged with respect to Card D, should be more distant (farther behind Card D) when Card B was present (Task 1, Fig. 2a) rather than absent (Task 2, Fig. 2b). Therefore, the obtained difference between Tasks 1 and 2 is a measure of the *effect of Card B* on the perceived depth position of Card A. Similarly, if the size cue between Cards A and D contributed to the perception of the depth position of Card A, the perceived depth position of Card A, as judged with respect to Card B, should be more distant (farther behind Card B) when Card D was absent (Task 3, Fig. 2c) rather than present (Task 4, Fig. 2a). Therefore, the obtained difference between Tasks 3 and 4 is a measure of the *effect of Card D* on the perceived depth position of Card A. According to the adjacency principle, the perceived position of Card A should be more affected by Card B than by Card D. The purpose of Task 5 was to determine that the interposition cue was effective in modifying the perceived depth

TABLE 1

Average Perceived Depth in Centimeters between Playing Cards as Measured by the Hand Adjustments

Presence or Absence of B	Perceived Depth of A behind D	Presence or Absence of D	Perceived Depth of A behind B
B Present	38.2**	D Absent	30.5**
B Absent	−0.3	D Present	27.0**
Diff.	38.5**	Diff.	3.5*

* $p < .05$ (t test).
** $p < .01$ (t test).

between B and D from that which normally would have been expected as a consequence of the size difference between these two cards.

In the instructions to *O* a small pegboard with two movable pegs was used to illustrate the difference between frontoparallel and depth extents, in order to be certain that *O* understood that the distance between the playing cards which was to be measured by the hand adjustment was the depth component only. Following a set (three) of hand adjustments to indicate the perceived depth between a specified pair of playing cards, *O* turned and faced the light-adaptation surface which was turned on for approximately 10 sec. This procedure of obtaining three measurements of apparent depth followed by a period of light adaptation was repeated until all five tasks had been completed.

Results

The average results from the hand adjustments are shown in Table 1. A positive value in the left-data column of Table 1 indicates that, on the average, Card A was perceived as being more distant than Card D. A negative value indicates the converse. A positive value in the right-data column of Table 1 indicates that, on the average, Card A was perceived as being more distant than Card B. One potential *O* who failed to perceive Card A as behind Card B in the situation in which Card D was absent (Task 3) was not used in the experiment. A positive response on Task 3 was considered to be essential to the experiment, since it indicated that *O* was responsive to the size cue to relative depth. The error terms used in the t tests in the study were computed from the distribution of 20 scores or 20 differences whose mean was to be tested for significant deviation from zero. The 38.5-cm. difference in Table 1 represents the effect of Card B upon the perceived depth position of Card A and the 3.5-cm. difference represents the effect of Card D upon the perceived depth position of Card A. The former effect is greater than the latter effect at the .01 level of confidence; $t\ (19) = 5.42$. As expected from the adjacency principle, the size cue between Playing Cards A and B was more effective than the size cue between Playing Cards A and D in determining the apparent position of Playing Card A.

The interposition cue on the average made Card B appear to be located 20.4 cm. behind Card D with this value different from zero at the .01 level of confidence; $t\ (19) = 3.68$.

Discussion

The results from this study indicate that the adjacency principle can be applied to the size cue to relative depth. The size cue between Card A and the adjacent Card B was dominant in determining the perceived depth position of Card A. This occurred even when *O*'s task was to judge the depth between Card A and the displaced Card D. It is clear that *O* cannot use the size cue to relative depth between one object and a displaced object independently of size cues between this object and adjacent objects.[3] However, the results also indicate that while the size cue between Playing Cards A and D was not very effective in determining the perceived position of A, it was not completely ineffective as shown by the statistical significance of the 3.5-cm. difference in Table 1. The present study supports the hypothesis that the perceived position of one object (Object A) in a configuration of objects is determined by whatever distance cues are present between Object A and other objects, with the effectiveness of such cues being inversely related to the relative separation of the objects from Object A. This is a statement of the adjacency principle.

---

[3] This does not eliminate the possibility, however, that the task set of *O* (to judge the depth between A and one of the other cards) has a limited effect upon the relative efficiency of the several size cues in determining the apparent position of A (cf. Gogel, 1963a).

In a previous study (Gogel & Harker, 1955) it had been found that the relative effectiveness of size cues increased with an increase in their lateral separation. It was later found (Gogel, 1956) that this result was attributable to a decrease in the effectiveness of the "equidistance tendency" as a function of the increased lateral separation, not to an increase in the effectiveness of the size cue to relative depth. Thus, the results from the present study and the previous study are compatible. Also, it is clear that the results of the present study cannot be explained by the "equidistance tendency" since in the present study Card A appeared behind, not equidistant with, Card B (cf. Gogel, 1963a).

The adjacency principle is in an experimentally formative stage. In order for the principle to represent a general phenomenon, it must apply to many cue systems of relative depth and to many apparent characteristics of objects in addition to that of relative depth. Furthermore, the question occurs as to whether adjacency should be defined in physical, retinal or perceptual terms (Gogel, 1963b). Although there is some evidence available concerning this question (Gogel, 1963a), a definitive answer requires further research. Finally, in order for the adjacency principle to be expressed quantitatively, the relation between adjacency and relative cue effectiveness must be measured for a variety of cue systems.

REFERENCES

GOGEL, W. C. The tendency to see objects as equidistant and its inverse relation to lateral separation. *Psychol. Monogr.*, 1956, **70**(4, Whole No. 411).

GOGEL, W. C. Size cues and the adjacency principle. *Civ. Aeromed. Res. Inst. Rep.*, 1963, No. 63-28. (a)

GOGEL, W. C. The visual perception of size and distance. *Vision Res.*, 1963, **3**, 101–120. (b)

GOGEL, W. C., & HARKER, G. S. The effectiveness of size cues to relative distance as a function of lateral visual separation. *J. exp. Psychol.*, 1955, **50**, 309–315.

(Received May 20, 1964)

ournal of Experimental Psychology
965, Vol. 70, No. 3, 294–299

EFFECT OF PAIRING A STIMULUS WITH PRESENTATIONS OF THE UCS ON THE EXTINCTION OF AN AVOIDANCE RESPONSE IN HUMANS[1]

ROBERT K. BANKS

University of Waterloo

Experimentation was carried out to study the effect of pairing a stimulus (PS) with UCS presentations, on the rate of extinction of an avoidance response in humans. It was found that extinction was facilitated by any manipulations of the PS which produced a difference between the acquisition and extinction situations. The greater the difference the more rapid was the extinction. The effect of the PS was interpreted through the generalization-decrement hypothesis and it was suggested that the effect of the PS would be replicable when the stimulus was not paired with presentations of the UCS.

The avoidance behavior of both animals and humans is generally accepted to be highly resistant to experimental extinction. While much of the literature supports this point of view (e.g., Logan, 1951; Sheffield & Temmer, 1950; Solomon, Kamin, & Wynne, 1953), results reported by Finch and Culler (1935) suggest that relatively rapid extinction can be achieved by a modification within the avoidance-conditioning paradigm. Finch and Culler, using a 1,000-cps tone as a CS and shock as a UCS, studied the extinction of a paw-withdrawal response in dogs under two experimental procedures. In both, an additional stimulus, a buzzer, was included in the conditioning paradigm and occurred simultaneously with all presentations of the avoidable shock. In one procedure this paired stimulus (PS) was withheld during extinction trials and in the other procedure the PS was presented during extinction trials. The two procedures can be characterized as follows: (*a*) the PS was presented on all acquisition trials and withheld on all extinction trials, a 100%-0% treatment, and (*b*) the PS was presented on all acquisition and all extinction trials, a 100%-100% treatment.

Finch and Culler (1935) found that resistance to extinction was typically high under their 100%-100% treatment, but extinction occurred more than three times as rapidly under their 100%-0% treatment. Although the addition of the PS during conditioning (100%-0%) appears to reduce the resistance to extinction of avoidance behavior, the failure of Finch and Culler to include a control treatment in which the PS was never presented (0%-0%) prohibits a firm conclusion. Their data are open to two interpretations; that the 100%-100% treatment increased resistance to extinction or that the 100%-0% treatment reduced resistance to extinction. This paper reports experiments designed to determine whether the addition of the PS to the avoidance paradigm facilitates extinction and to explore the parameters of this effect.

---

[1] This paper is based on a dissertation submitted to the School of Graduate Studies f the University of Toronto in partial fulfillment of the requirements for the PhD degree n psychology. The author is indebted to Glen E. MacDonald, George Mandler, and Richard H. Walters who served as members f his dissertation committee.

EXPERIMENT I

In order to determine whether the addition of the PS to the avoidance paradigm facilitates extinction of avoidance behavior, response strength during extinction was compared under a 100%-0% and a 0%-0%, or control, treatment.

Method

Subjects.—Twenty undergraduates at the University of Toronto served as *S*s. They were randomly divided into two groups of equal size.

Groups.—Two groups (100%–0% and 0%–0%), differing only in PS frequency during acquisition trials, were employed.

Apparatus.—The avoidance-conditioning situation was designed to resemble that used by Finch and Culler and was constructed for conditioning a hand-withdrawal response. A grid with eight bars attached alternately to the plus and minus terminals of an electric pulse generator was mounted on top of a hollow wooden box. The grid was hinged to one end of the box and a spring held the grid in a raised, or open, position. The weight of a hand was sufficient to close the grid. A microswitch was fastened to the rim of the box so that when the grid was closed the switch was depressed. A signal light was connected to the microswitch.

The CS was the light from a 60-w. light bulb placed directly under the grid in the bottom of the box. The UCS was an electric shock with a potential of 125 v. dc at the grid. The PS, the stimulus paired with shock presentation, was the sound of a 5-v. door buzzer. Hunter interval timers regulated the onset and offset of the CS, UCS, and PS.

Procedure.—The *S* was seated beside the grid and was instructed to place his hand on it. He was told that the experiment consisted of a number of trials and that at the beginning of each he was to rest his hand on the grid but that he did not have to leave his hand on the grid throughout the trial. He was then informed that at times he would receive an electric shock but that as soon as he learned how to avoid the shock he could do so as often as he wished.

The duration of the CS was 1 sec. The CS was terminated by the onset of the UCS, the duration of which was also 1 sec. When the PS was presented, its onset and termination coincided with those of the UCS.

A response was counted as a CR if *S* removed his hand from the grid before the electric shock was delivered. The *E* could determine whether the hand had been removed by watching the signal connected to the microswitch in the box. Before acquisition trials were begun, the CS was presented twice while *S*'s hand rested on the grid. If *S* lifted his hand on either occasion, presentations of the CS were continued until *S* failed to respond on two consecutive occasions.

Eighty acquisition trials were administered. The UCS was presented on each of these trials but was avoided by *S* if the CR occurred prior to the onset of the UCS. Forty extinction trials, in which the UCS and PS were withheld, were administered to each *S*. Five intertrial intervals, 2, 4, 6, 8, and 10 sec., were randomly assigned to acquisition and extinction trials and were the same for all *S*s.

Results

During acquisition trials the 100%-0% group gave a mean of 3.4 nonavoidance responses compared to 3.1 for the 0%-0% group, indicating that PS frequency had little effect on response strength during acquisition. During extinction, however, the 100%-0% group gave a mean of 24.4 CRs ($SD = 12.2$) compared to 36.3 ($SD = 10.1$) for the 0%-0% group, $t\ (18) = 2.37$, ($p < .05$). The results therefore indicate that the addition of a PS to the avoidance paradigm reliably facilitates the extinction of an avoidance response.

EXPERIMENT II

In Exp. I the two groups differed from each other, not only in the frequency of PS presentations during acquisition, but automatically, in the amount of difference in PS frequency between the acquisition and extinction periods. The 100%-0% treatment produced a difference of 100% in PS frequency while the control treatment (0%-0%) produced no difference in PS frequency. This observation suggested the hypothesis, explored in

Exp. II, that the operative role of the PS in facilitating extinction is that of creating a difference in PS frequency between the acquisition and extinction situations; the greater the difference, the more rapid the extinction.

Studies of the extinction of non-avoidance behavior lend support to this hypothesis. The relevant literature indicates that extinction is facilitated by increasing differences between acquisition and extinction situations in a number of variables, e.g., amount of reinforcement (Lewis & Duncan, 1957), frequency of reinforcement (Lewis & Duncan, 1957), intertrial interval (Reynolds, 1945), and deprivation level (Yamaguchi, 1952). It is possible that the rate of extinction of avoidance behavior is similarly sensitive to differences between acquisition and extinction.

The hypothesis that the operative role of the PS is that of creating differences between acquisition and extinction has an important implication. It follows that the resistance to extinction of an avoidance response should be reduced, not only by the PS manipulation used in Exp. I but by any PS manipulations which produce a difference in PS frequency between acquisition and extinction.

Differences between acquisition and extinction can be produced by a large number of PS manipulations. For example, a difference of 100% can be produced by a 100%-0% treatment, as in the first experiment, or by a 0%-100% treatment. A difference of 0% can be produced by a 0%-0% treatment as before, or by a 1%-1% treatment, a 2%-2% treatment, etc., up to a 100%-100% treatment. The same is true for intermediate amounts of difference. For example, a difference of 50% can be produced by a 50%-0% treatment, a 51%-1% treatment, a 52%-2% treatment, etc., up to a 100%-50% treatment or by a 0%-50% treatment, a 1%-51% treatment, etc., up to a 50%-100% treatment. An examination of the alternate treatments by which any given amount of difference can be produced shows that they differ from each other, along one or both of two dimensions. The first is the overall frequency of PS presentations (e.g., 0%-0% vs. 100%-100%) and the second is the direction of change in frequency from acquisition to extinction (e.g., 100%-0% vs. 0%-100%). If, as hypothesized, the PS has its effect by producing differences in PS frequency between acquisition and extinction it follows that resistance to extinction should decrease as the amount of difference increases when the difference is produced by (*a*) presenting the PS with its maximum as well as its minimum overall frequency, and by (*b*) increasing as well as decreasing PS frequency from acquisition to extinction. Experiment II was carried out to test these predictions.

Method

Subjects.—One hundred and forty-four undergraduates at the University of Toronto served as *S*s. They were randomly assigned to 12 groups of equal size.

Groups.—The effect of the PS under maximum and minimum overall PS frequencies was examined at the 0%, 50%, and 75% levels of difference in PS frequency. At the 0% level the maximum frequency was represented by a 100%-100% treatment and the minimum by a 0%-0% treatment. At the 50% level the maximum frequency was represented by a 100%-50% and a 50%-100% treatment and the minimum by a 50%-0%, and a 0%-50% treatment. At the 75% level the maximum was represented by a 100%-25%, and a 25%-100% treatment and the minimum by a 75%-0%, and a 0%-75% treatment. The 100% level of difference was omitted because such a difference can only be produced by a 100%-0% or a 0%-100% treatment, both of which require the same overall frequency of PS presentation.

The effect of the PS under increases and decreases in PS frequency from acquisition to extinction was examined at the 100%, 75%, and 50% levels of difference in PS frequency. At the 100% level increases were represented by a 0%-100% treatment and decreases by a 100%-0% treatment. At the 75% level increases were represented by the 0%-75% and 25%-100% treatments and decreases by the 100%-25% and 75%-0% treatments. At the 50% level increases were represented by the 0%-50% and 50%-100% treatments and decreases by the 50%-0% and 100%-50% treatments. The 0% level was omitted because it involves neither an increase nor a decrease in PS frequency from acquisition to extinction. The 25% level was omitted throughout the experiment because it was felt that testing at the 100%, 75%, 50%, and 0% levels of difference represented a sufficiently broad sample of the range of differences in PS frequency.

Apparatus and procedure.—The apparatus was identical to that used in Exp. I. The procedure differed in that 60 acquisition and 60 extinction trials were administered to all *Ss*. This was done so that the various PS manipulations at each level of difference produced equal absolute as well as equal percentage differences in PS frequency. When the PS was presented on less than 100% of the acquisition or extinction trials it was presented on a different random schedule for each *S*. When presented during extinction, the PS filled the interval in which the UCS would have occurred.

Results

Gradients of response strength during extinction under both maximum and minimum frequencies of PS presentation and under both increases and decreases in PS frequency from acquisition to extinction are presented in Fig. 1. Under both maximum and minimum PS frequencies and under both increases and decreases in PS frequency from acquisition to extinction, resistance to extinction consistently varied as an inverse function of the amount of difference in PS frequency. Inspection of Fig. 1 also shows a pronounced tendency for frequency decreases to have a stronger effect than frequency increases and a slight tendency for minimum overall frequencies to have a stronger effect than maximum frequencies.

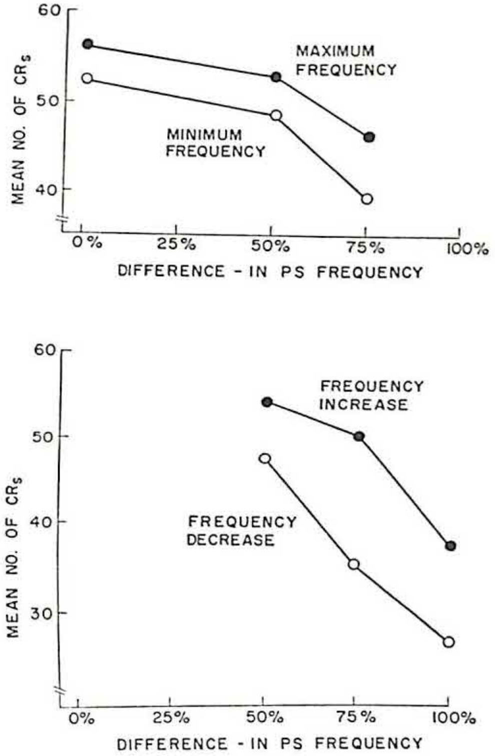


FIG. 1. Mean number of CRs in 60 extinction trials as a function of amount of difference in PS frequency under maximum and minimum PS frequencies and under increases and decreases in PS frequency.

A two-way variance analysis of the number of CRs given under maximum and minimum frequencies at the 0%, 50%, and 75% levels of difference yielded an F (2, 114) $= 3.76$, $p < .05$ for the effect of the amount of difference in PS frequency, and failed to indicate a significant difference between the effects of maximum and minimum PS frequencies ($p > .20$). A similar analysis of the number of CRs given under increases and decreases in PS frequency at the 100%, 75%, and 50% levels of difference yielded an F (2, 114) $= 6.33$, $p < .005$ for the effect of the amount of difference and an $F(1, 114) = 7.71, p < .01$ for the effect of the direction of change in PS frequency. The interaction effects failed to approach significance in both analyses.

In summary, the results indicate a reliable and consistent relationship between resistance to extinction and the amount of difference in PS frequency from acquisition to extinction. The results also show that the effect of the amount of difference in PS frequency is greater under decreases in PS frequency than under equal amounts of increase.

DISCUSSION

The data show that the typically high resistance to extinction of an avoidance response can be reduced in a marked and orderly fashion by varying PS frequency from acquisition to extinction. Probably the simplest interpretation of the effect of the PS is through the discrimination hypothesis (Mowrer & Jones, 1945) which holds that extinction is facilitated by any circumstance (e.g., differences in PS frequency) which makes it easy for *S* to perceive differences between the acquisition and extinction periods. A more complex but perhaps more fruitful interpretation of the PS can be attempted through the generalization-decrement hypothesis employed by Sheffield (1949) to account for the partial-reinforcement effect. From this point of view the effect of the PS would be attributed to traces of PS presentations persisting through time and forming part of the CS on subsequent trials. Increasing the difference in PS frequency from acquisition to extinction would thus decrease the degree to which the CS situation, typical of acquisition, was represented during extinction and so would produce more rapid extinction. Following this interpretation, the higher resistance to extinction under frequency increases than under decreases can be seen as an instance of response strength varying as a function of the magnitude of the CS situation. This is a phenomenon well documented in the literature (e.g., Grice & Saltz, 1950; Hovland, 1937; Spiker, 1956) and usually interpreted in terms of Hull's (1949) stimulus-intensity-dynamism postulate.

The apparent applicability of the generalization-decrement interpretation of the effect of the PS is of interest in light of the implicit assumption throughout the experiments that the pairing of the PS and the UCS was crucial to the experimental effect. If the PS (a buzzer) had its effect through its traces forming part of the CS situation of the following trial, as assumed by the generalization-decrement interpretation, the fact that the buzzer had been paired with the UCS was merely incidental, not crucial, to the effect. Any other positioning at the end of the trial would have sufficed. Moreover, because the intertrial interval varied between 2 and 10 sec., pairing the buzzer with the UCS resulted in the buzzer's preceding the onset of the next trial by a variable interval. As conditioning is better under a constant than under a variable interval, it can be suggested that pairing the buzzer with the UCS actually interfered with the potential experimental effect. A more powerful effect would be expected if the buzzer were presented so that it always preceded the next trial by a constant interval.

REFERENCES

FINCH, G., & CULLER, E. Relation of forgetting to experimental extinction. *Amer. J. Psychol.*, 1935, **47**, 656–662.

GRICE, G. R., & SALTZ, E. The generalization of an instrumental response to stimuli varying in the size dimension. *J. exp. Psychol.*, 1950, **40**, 702–707.

HOVLAND, C. I. The generalization of conditioned responses. II. The sensory generalization of conditioned responses with varying intensities of tone. *J. genet. Psychol.*, 1937, **51**, 279–291.

HULL, C. L. Stimulus intensity dynamism (*V*) and stimulus generalization. *Psychol. Rev.*, 1949, **56**, 67–76.

LEWIS, D. J., & DUNCAN, D. P. Expectation and resistance to extinction of a lever pulling response as functions of percentage of reinforcement and amount of reward. *J. exp. Psychol.*, 1957, **54**, 115–121.

Logan, F. A. A comparison of avoidance and non-avoidance eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 390–393.

Mowrer, O. H., & Jones, H. M. Habit strength as a function of the pattern of reinforcement. *J. exp. Psychol.*, 1945, **35**, 292–311.

Reynolds, B. Extinction of trace conditioned responses as a function of the spacing of trials during acquisition and extinction. *J. exp. Psychol.*, 1945, **35**, 81–95.

Sheffield, F. D., & Temmer, H. W. Relative resistance to extinction of escape training and avoidance training. *J. exp. Psychol.*, 1950, **40**, 287–298.

Sheffield, V. F. Extinction as a function of partial reinforcement and distribution of practice. *J. exp. Psychol.*, 1949, **39**, 511–526.

Solomon, R. L., Kamin, L. J., & Wynne, L. C. Traumatic avoidance learning: The outcomes of several extinction procedures with dogs. *J. abnorm. soc. Psychol.*, 1953, **48**, 291–302.

Spiker, C. C. The stimulus generalization gradient as a function of the intensity of stimulus lights. *Child Develpm.*, 1956, **27**, 85–98.

Yamaguchi, H. G. Gradients of drive stimulus (S_D) intensity generalization. *J. exp. Psychol.*, 1952, **43**, 298–304.

(Received May 23, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 3, 300–303

UNAVAILABILITY AND ASSOCIATIVE LOSS IN RI AND PI[1]

JOHN CERASO AND ANN HENDERSON

Yeshiva University

An experiment was reported in which 1st-list recall, 2nd-list recall, and single-list recall were measured shortly after learning, and after 24 hr. 3 measures of retention were employed: the aided-recall measure; a measure of item availability; a measure of associative pairing. The most striking result was that while 1st-list recall remained stable over 24 hr. as measured by aided recall, the availability measure showed increased recall, and the associative measure showed decreased recall. It was argued that this finding indicates that item recall and associative recall are 2 independent processes. The diminishing recall for the 2nd list over time seemed to be due to both item loss and associative loss. Recall at 24 hr. showed no difference between original list (OL) and interpolated list (IL) on any of the measures employed.

The notion that verbal learning involves two logically distinct aspects, item learning and associative learning, has recently been gaining wide acceptance (Rock & Ceraso, 1964; Underwood & Schulz, 1960). There have also been some attempts to extend this notion to problems of retention. For example, Rock and Ceraso argue that the widely used retention measure of aided recall may be affected by either item loss or associative loss, and that adequate theories of forgetting may be impossible without recognizing that these two types of forgetting can occur.

It is the purpose of this study to measure the associative and item loss occurring in retroactive inhibition (RI) and proactive inhibition (PI) and to chart any temporal changes occurring for these two factors.

METHOD

Procedure.—The experiments employed the method of modified free recall (MMFR) technique (Barnes & Underwood, 1959) to measure both RI and PI simultaneously. Four groups of *S*s were employed: two experimental groups, both of which learned two successive lists of the A-B, A-C paradigm, and two control groups, both of which learned only one list.

One experimental group was tested for recall 2 min. following interpolated learning, the other group was tested 24 hr. after the learning session. One control group was tested 2 min. after original list (OL), the other 24 hr. after learning. The 2-min. recall group will be called the immediate (Imm.) group.

The *S*s learned the lists in a memory drum using standard paired-associate techniques. Items were presented at a 2:2-sec. rate with a 4-sec. intertrial interval. Lists were learned to one perfect recitation. For the experimental *S*s a 2-min. interval was employed between the end of OL and the beginning of interpolated list (IL) learning. For the experimental group the recall test consisted of presenting each stimulus and asking *S* for both responses to the stimulus. The *S* was given up to 15 sec. to recall. After all 10 stimuli had been presented, *S* was asked to give any responses he might remember from the first or second list even if he did not recall the stimulus to which they had been learned. This test (free recall) was aimed at eliciting any "free-floating" responses which were available, but not paired with a stimulus. Finally, *S* was presented with the words he had recalled and was asked to indicate the list (1 or 2) in which they had been learned. When *S* originally recalled the words, a number from 1 to 20 was assigned to each word. When he was asked to assign the words to lists, the numbers were randomized and the words read in the randomized order.

---

[1] This work was supported by a grant to the senior author from the National Science Foundation (GB-1604).

Material.—The lists to be learned consisted of 10 pairs of nonsense syllables as stimuli coupled with four or five letter, one-syllable nouns as responses. The stimuli were taken from the Archer (1960) list and were 50 in association level. The nouns were A-level frequency nouns as given in the Thorndike-Lorge (1944) general count. Two separate lists (A and B) were constructed, having the stimulus items in common, but differing as to response items. The responses from A and B were equal in Thorndike-Lorge frequency of occurrence. Half the *S*s in each experimental group had A as a first list and B as a second list; the remaining half had the reverse. Half the control *S*s learned only A, half only B.

Subjects.—The *S*s were 64 students recruited at Pratt Institute, 16 *S*s in each group. The *S*s were randomly assigned to the four groups. If an *S* did not reach criterion on the first list within 20 trials, he was dismissed from the experiment. Roughly one third of the *S*s were lost because of failure to reach criterion.

RESULTS

Table 1 presents the mean learning trials for the four groups. An *F* test revealed no significant learning differences between the groups. The major results of the experiment are presented in Table 2. This table presents a complete analysis of the "fate" of OL, IL, and control list (CL) responses. The categories employed were: (Category 1) total number of responses given to the appropriate stimulus and either (*a*) correctly or (*b*) incorrectly assigned to list; (Category 2) total number of responses given to an inappropriate stimulus and either (*a*) correctly or (*b*) incorrectly assigned to list; (Category 3) total number of responses given in the free-recall test; (Category 4) total number of responses never appearing in recall. A response was placed in Category 3 only if it had not been given in Category 1 or 2. A response was placed in Category 2 only if it had not appeared in Category 1. As can be seen from Table 2 the number of responses accounted for by this scheme exhausts all responses.

TABLE 1

MEAN TRIALS TO LEARN THE FIRST LIST

Group	Mean Trials
Exp. Imm.	11.88
24 hr.	12.13
Control Imm.	13.38
24 hr.	13.50

Category 1*a* in Table 2 represents the requirements imposed upon recall in an aided-recall test. The *S* must be able to produce the response, he must produce it to the appropriate stimulus, and he must know the list membership of the response, since *E* usually asks him to produce the response from a given list. Therefore, the number of responses in 1*a* will be taken as the measure of aided recall. Two analyses of variance were performed on this measure, one comparing OL with CL, the other comparing

TABLE 2

"FATE" OF ORIGINAL-LIST, INTERPOLATED-LIST, AND CONTROL-LIST RESPONSES

List	+Stimulus		−Stimulus		Category 3	Category 4
	(1*a*) +List	(1*b*) −List	(2*a*) +List	(2*b*) −List	Free Recall	Not Given
OL Imm.	94	3	0	1	8	54
24 hr.	95	11	7	3	6	38
IL Imm.	149	2	1	0	4	4
24 hr.	111	9	1	3	7	29
CL Imm.	155		2		0	3
24 hr.	151		1		1	7

Journal of Experimental Psychology
1965, Vol. 70, No. 3, 304–308

ACQUISITION-TRANSFER AND RETENTION OF S-R$_1$, S-R$_2$ ASSOCIATIONS AS A FUNCTION OF S-R$_1$, S-R$_2$ PATTERN[1]

RICHARD POPP AND JAMES F. VOSS

University of Pittsburgh

6 S-R$_1$, S-R$_2$ modified paired associates, each with P(R$_1$) = P(R$_2$) = .5, were presented for 40 trials under 5 response-pattern (RP) conditions: random, and alternation on each block of 1, 5, 10, or 20 trials. Recall and recognition tests of "unlearning" were given. The results indicated: (a) anticipations of either R$_1$ or R$_2$ did not vary significantly with RP; (b) correct responses increased as RP varied from the random to 20-trial conditions; (c) successive R$_1$-R$_2$ shifts diminished transfer effects; (d) transfer varied from negative to positive as a function of number of R$_1$ trials before the 1st shift to R$_2$; (e) "unlearning" decreased from the 20-trial to the single-alternation conditions for the recall measure. "Unlearning" was not found with the recognition measure.

Voss, Thompson, and Keegan (1959) employed a modified S-R$_1$, S-R$_2$ paired-associate procedure in which the probability ratio of the randomly ordered responses was varied from .5-.5 to 1.0-0.0. It may be noted that the .5-.5 ratio of this design may be considered as an A-B, A-C transfer situation with random B and C occurrence. Similarly, the A-B, A-C transfer design in which A-B and A-C presentations are equal may be considered as a probabilistic situation with a .5-.5 response probability ratio and successive B and C occurrence. Thus, the major difference of the response probability and the transfer designs is response pattern.

The present experiment was designed to provide more information regarding the role of response pattern in S-R$_1$, S-R$_2$ acquisition-transfer performance. The use of five response-pattern conditions and modified free-recall and recognition tests provided for the study of response pattern upon acquisition, specific A-B, A-C transfer effects, and "unlearning" (Barnes & Underwood, 1959).

METHOD

Procedure.—A list of six S-R$_1$, S-R$_2$ paired associates was presented by the anticipation method (S, S-R$_1$ or S, S-R$_2$) for 40 trials in each of five experimental conditions. The $P(\text{S-R}_1) = P(\text{S-R}_2) = .5$, i.e., R$_1$ and R$_2$ each occurred for 20 trials. The independent variable consisted of five R$_1$-R$_2$ response-pattern conditions (RP). (R$_1$ denotes the first response and R$_2$ denotes the second response presented with a particular stimulus.) Condition R: random presentation of R$_1$ and R$_2$. The random order of R$_1$ and R$_2$ was different for each of the six associations. The mean number of trials on which alternations occurred was 19.3 and the average trial-run length for R$_1$ and R$_2$ combined was 1.9 trials. Condition 1: single alternation in which R$_1$ and R$_2$ alternated on each trial. Condition 5: R$_1$ was presented on Trials 1–5, R$_2$ on Trials 6–10, R$_1$ on Trials 11–15 . . . R$_2$ on Trials 36–40. Condition 10: R$_1$ on Trials 1–10, R$_2$ on Trials 11–20, R$_1$ on Trials 21–30, and R$_2$ on Trials 31–40. Condition 20: R$_1$ on Trials 1–20, and R$_2$ on Trials 21–40.

The modified paired associates were presented in a different random order on each trial at a 2:2-sec. rate with a 2-sec. intertrial interval. The instructions were those of Voss, Thompson, and Keegan (1959).

Following completion of the 40 trials, *S*s immediately were given the recall test. Each *S* was presented one of the stimulus items on a 5 × 7 in. card and asked to recall the last

---

[1] This experiment was supported by the National Institute of Mental Health (M-3531) and Child Health and Human Development (HD 00957-01).

TABLE 1

MEAN ANTICIPATIONS, CORRECT RESPONSES, R_1 CORRECT RESPONSES, AND R_2 CORRECT RESPONSES FOR THE RESPONSE-PATTERN CONDITIONS

RP Cond.	$R_1 + R_2$ Antic.	Correct Responses	Correct R_1 Responses	Correct R_2 Responses
R	129.8	62.4	35.2	27.3
1	126.2	77.0	43.3	33.7
5	111.8	99.4	53.0	46.4
10	147.3	140.6	71.6	69.0
20	152.8	152.5	73.1	79.4

response item that had been presented with that stimulus item. Following a response or the period of 10 sec., *S* was asked to recall the other response item that had been presented with that stimulus.[2] The remaining five stimuli were presented in the same manner. In no case was *S* given any information by *E* regarding correct responses. The order of presentation of the six stimuli was varied by use of a randomly permuted 6 × 6 Latin square with order of presentation orthogonal to experimental condition.

Following completion of the recall task, each *S* immediately was presented with a sheet which provided instructions to match the 12 responses presented on the right side with the six stimuli on the left side of the sheet. Both stimuli and responses were listed on the sheet in random order. After *E* read the instructions and *S* followed by referring to his sheet, *S* wrote the numbers denoting the response items in two boxes provided with each stimulus item. There was a 1-min. time limit.

Subjects.—The *S*s were volunteers from the introductory psychology class at the University of Pittsburgh. The *N* was 60 with an $N = 12$ for each experimental condition. The *S*s were assigned to the experimental conditions on the basis of a table of random numbers.

Verbal materials.—The verbal material consisted of low-frequency words employed by Horton and Kjeldergaard (1961). The specific S-R_1, S-R_2 pairings were those employed previously by Morning and Voss (1964).

The words employed as R_1 and R_2 were counterbalanced in each RP condition so that one half of the *S*s in each condition received, e.g., CAIRN as R_1 to the stimulus MERLE and WINCH as R_2; for the other one half of the *S*s, WINCH was R_1 and CAIRN was R_2.

RESULTS

$R_1 + R_2$ anticipations.—The first column of Table 1 presents the mean $R_1 + R_2$ anticipation frequency as a function of RP. It should be noted that this measure involves an R_1 or an R_2 anticipation for a particular S-R_1, S-R_2 association, regardless of whether R_1 or R_2 occurred on that specific trial. An analysis of variance of the $R_1 + R_2$ anticipation data revealed that the RP source of variation is not significant, $F(4, 55) = 2.04$, $p > .05$.

$R_1 + R_2$ correct responses.—Table 1 also presents mean correct responses as a function of RP. An analysis of variance performed upon the correct-response data revealed that the RP source of variation is significant, $F(4, 55) = 16.01$, $p < .01$. Thus, correct responses increased as RP varied over Cond. R, 1, 5, 10, and 20.

Trial data.—Figure 1 presents the mean $R_1 + R_2$ correct responses as a function of trial for the RP conditions. In Cond. 5, 10, and 20, the shifts in correct responses are readily apparent with the shift in R_1-R_2 presentation. Condition 1 yielded a saw-toothed effect in early trials with correct responses greater for R_1. It should be noted that the inspection of the data of Cond. 1 also indicates that not one *S* "caught on" to the single

---

[2] In Cond. 1, 5, 10, and 20, the "last" response was always R_2 and "other" response was always R_1. However, in Cond. R, the Trial 40 response was R_1 for two of the six associations. Hence, for recall and recognition, the Trial 40 response is termed "last" response and the second response is termed "other" response.

IL with CL. These analyses were followed by comparison of the means using the Newman-Keuls technique described in Winer (1962). The results of these analyses were: Recall for OL was significantly poorer than recall for CL ($p < .01$), for both Imm. and 24-hr. tests. There were no significant OL or CL changes over 24 hr. Recall for IL was not significantly different from CL immediately, but was different after 24 hr. ($p < .01$). The IL Imm. vs. IL 24-hr. recall difference was also significant ($p < .01$).

Recall for OL and IL were directly compared using t tests for related samples (since the same *S*s are tested for both lists). The results of these analyses were a significant OL-IL difference ($p < .05$) when recall is tested immediately, but no significant difference at the 24-hr. recall.

The results of these analyses of the aided-recall measure are quite consistent with those typically reported (Slamecka & Ceraso, 1960) which show a relatively stable level of recall for the first list coupled with a decreasing level of recall for the second list. That PI may increase over time to a point where it is equal to RI is also found in the literature (e.g., Koppenaal, 1963).

The next analysis of the data concerned the number of response items available, irrespective of whether these responses were correctly paired or not. The total number of items available is obtained from Table 2 by summarizing all categories except the Not Given category (or by subtracting the Not Given category from 160). Again, two analyses of variance were performed, each followed by multiple comparisons of the means. The result of the OL vs. CL analysis was: a significant *rise* in OL response available over time, and significant differences between OL and CL available responses at both the Imm. and 24-hr. tests. The differences are all significant at $p < .01$. There was no difference in the availability of CL responses over time. The result of the IL vs. CL analysis was: a significant *decrease* in IL responses available over time, and a significant difference between IL and CL responses only at the 24-hr. recall, these differences being significant at $p < .01$. The number of available responses for OL and IL were directly compared. The results of these analyses were a significant OL-IL difference ($p < .01$) only for the Imm. test.

The results of these analyses of the availability measure show striking differences in the temporal changes occurring for OL vs. IL lists. The OL list shows increasing availability over time, the IL list shows decreasing availability.

The aided-recall measure previously analyzed was used as the basis for analyzing *S*'s ability to produce the correct associate (to both stimulus and list) under the various conditions. However, since the number of responses correctly associated could be a function of the number of responses available, a somewhat different measure of association was employed. For each *S* the number of items correctly associated (Category 1*a*) was expressed as a percentage of the total number of items available. The mean of these percentages for the various conditions are presented in Table 3.

TABLE 3

MEAN PERCENTAGE OF CORRECT ASSOCIATION FOR AVAILABLE RESPONSES

List	Imm. Test	24-Hr. Test
OL	89.50	77.19
IL	95.38	85.56
CL	98.75	98.38

The analyses of variance for the association measure revealed the following: The OL differed significantly from the CL at both the Imm. ($p < .05$) and the 24-hr. ($p < .01$) recall test. There was no significant change over time in ability to associate for the CL. The IL did not differ significantly from the CL immediately but did differ at 24 hr. ($p < .01$). The IL showed a significant decrease in the associative measure over 24 hr. ($p < .01$). The direct comparison of OL and IL showed no significant differences at either the Imm. or 24-hr. test.

By this measure of association both OL and IL suffer associative loss over time as compared to CL.

Discussion

The three measures of recall used in the present experiment each give a different impression of the temporal course of first-list retention. While the aided-recall measure shows virtually no change over 24 hr., the availability measure shows increased recall, the association measure shows decreased recall. It is tempting to ascribe the "no change" in the aided-recall measure to the joint action of two opposed processes: an increase in the number of first-list items available over time, coupled with a growing difficulty in the correct assignment of these items to stimulus or to list.

The data are, however, subject to another interpretation. It could be that IL learning disrupts both the availability and associative properties of some OL items and that over time these items become more available, but do not recover their associative properties. This sequence of events would also give rise to an increased availability score and a decreased associative score as they have been used in the present analysis. The implication of this second interpretation would be that there is no growing associative loss for the OL over time, but only an increase in the number of items available whose associative properties were already lost during IL.

Either of these interpretations, however, assumes that over time the availability of an item can change in a way different from the change occurring for the associative properties of that item. Thus, the findings give support to the notion that associative loss and item loss are to some extent independent processes.

Recall for the second list over the 24-hr. interval declines for each of the measures used. The simplest explanation for these results is that PI is characterized by both associative loss as well as loss of availability over time.

It is interesting to note that at 24 hr., recall of the OL and the IL does not differ for any of the three measures employed. This suggests that the processes responsible for differences between RI and PI up to this point are transitory, and that after a reasonable period of time OL exerts the same kind and degree of interference on IL as IL does on OL.

REFERENCES

Archer, E. J. A re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychol. Monogr.*, 1960, **74**(10, Whole No. 497).

Barnes, J. M., & Underwood, B. J. "Fate" of first-list associations in transfer theory. *J. exp. Psychol.*, 1959, **58**, 97–105.

Koppenaal, R. J. Time changes in the strengths of A-B, A-C lists: Spontaneous recovery? *J. verbal Learn. verbal Behav.*, 1963, **2**, 310–319.

Rock, I., & Ceraso, J. Toward a cognitive theory of associative learning. In C. Scheerer (Ed.), *Cognition: Theory, research, promise.* New York: Harper & Row, 1964.

Slamecka, N. J., & Ceraso, J. Retroactive and proactive inhibition of verbal learning. *Psychol. Bull.*, 1960, **57**, 449–457.

Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, 1944.

Underwood, B. J., & Schulz, R. W. *Meaningfulness and verbal learning.* Chicago: Lippincott, 1960.

Winer, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

(Received May 27, 1964)

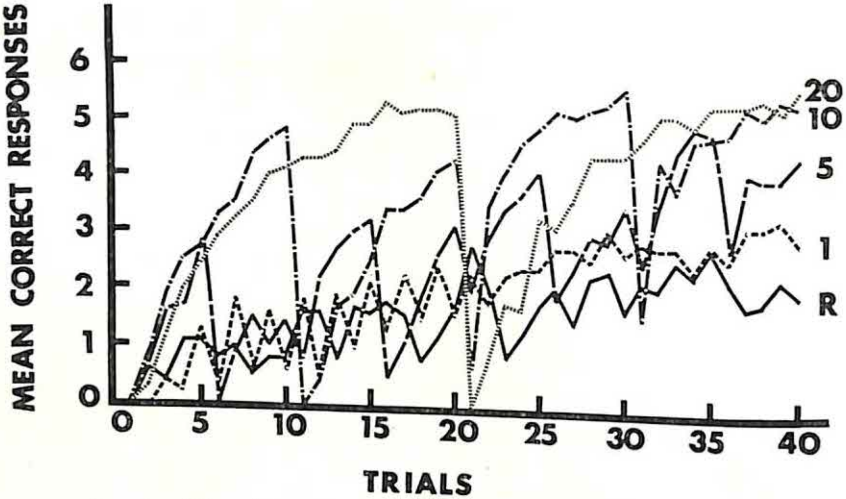


FIG. 1. Mean $R_1 + R_2$ correct responses as a function of trial with response pattern (RP) as parameter.

alternation pattern of each association. An analysis of variance was not performed on these data because the difference of RP conditions over trials was obvious. Although not presented, the $R_1 + R_2$ anticipation data yielded similar effects.

R_1-R_2 data.—The mean correct R_1 and R_2 responses for each RP condition are presented in Table 1. An analysis of variance revealed a greater frequency of R_1 than R_2 correct responses, $F(1, 55) = 7.97$, $p < .01$. In addition, a significant R_1 vs. R_2 $\times$ RP interaction, $F(4, 55) = 3.88$, $p < .01$, indicates that the difference of R_1 and R_2 correct responses decreased over Cond. 1, 5, and 10 with more correct R_2 than R_1 responses in Cond. 20. Identical analyses performed on the R_1 and R_2 correct-response data after five presentations of R_1 and R_2 (5-trial data) and after 10 presentations of R_1 and R_2 (10-trial data) revealed that the same significant sources of variation and systematic R_1 vs. R_2 changes over RP occurred regardless of stage of learning. Analyses of the anticipation data identical to those performed upon the correct-response data revealed that the same significant effects occurred for the anticipation data.

TABLE 2

PERCENTAGE OF TRANSFER AS A FUNCTION OF RP FOR THE 5-TRIAL, 10-TRIAL, AND 20-TRIAL DATA

RP Cond.	5 Trial	10 Trial	20 Trial
R	−20.0	−12.8	−12.7
1	−50.0	−34.3	−12.6
5	−15.9	−13.9	−6.6
10	−6.1	−6.4	−1.8
20	+11.7	+7.5	+4.2

Transfer.—The following measure of the transfer effect of R_1 upon R_2 was employed: [(No. Correct R_2 Responses − No. Correct R_1 Responses) ÷ (No. Correct R_2 Responses + No. Correct R_1 Responses)] $\times$ 100. Table 2 presents this measure for the 5-trial, 10-trial, and 20-trial data of the RP conditions. The following results may be noted: (*a*) negative transfer decreases over Cond. 1, 5, and 10; (*b*) in Cond. 20, transfer is positive; (*c*) transfer effects tend to decrease as a function of stage of training.

The 5-trial data of Cond. 5, 10-trial data of Cond. 10, and 20-trial data of Cond. 20 provide a measure of transfer as a function of stage of training after

TABLE 3

PERCENTAGE OF RESPONSES RECALLED CORRECTLY, PERCENTAGE OF "LAST" AND "OTHER" RESPONSES RECALLED CORRECTLY, PERCENTAGE OF RESPONSES RECOGNIZED CORRECTLY, AND PERCENTAGE OF "LAST" AND "OTHER" RESPONSES RECOGNIZED CORRECTLY AS A FUNCTION OF RP CONDITION

Exp. Cond.	Percentage					
	Recalled Correctly	"Last" Recalled Correctly	"Other" Recalled Correctly	Recognized Correctly	"Last" Recognized Correctly	"Other" Recognized Correctly
R	40.3	44.4	36.1	84.0	79.2	88.9
1	33.3	31.9	34.7	85.4	87.5	83.3
5	72.2	77.8	66.7	85.4	87.5	83.3
10	81.3	93.1	69.4	92.4	94.4	90.3
20	53.5	86.1	20.8	88.2	91.7	84.7

only one R_1-R_2 shift. The respective -15.9%, -6.4%, and 4.2% findings indicate that as stage of training increases, the first shift of R_1 to R_2 produces decreasing negative to positive transfer. It should be noted that the extremely large correct-response occurrence of R_1 compared to R_2 in Cond. 1 is in agreement with these data.

Another way of viewing transfer within the present design is to determine the effect of successive shifts upon transfer within a particular RP condition. Using the same transfer measure, the percentage of transfer generally decreased as a function of successive R_1-R_2 shifts for Cond. 1, 5, 10, and 20. For Cond. 10, percentage of transfer of the two successive R_2-R_1 shifts was -5.7 and 0.9. For Cond. 5, percentage of transfer was -15.9, -12.7, -7.8, and -1.9 for the successive R_2-R_1 shifts. For Cond. 1, percentage of transfer changed from a maximum of -57.9 to 7.9.

Recall and recognition.—Table 3 presents percentage of responses correctly recalled and percentage of responses correctly recognized as a function of RP. The table also presents the percentage correctly recalled and recognized of the "last" response presented (Trial 40) and the percentage correctly recalled and recognized of the "other" response.

An analysis of variance performed upon the recall frequency data revealed a significant RP source of variation, $F\ (4, 55) = 10.29$, $p < .01$, significantly more "last" responses recalled than "other" responses recalled correctly, $F\ (1, 55) = 65.75$, $p < .01$, and a significant interaction of the two main effects, $F\ (4, 55) = 20.89$, $p < .01$. The significant interaction is attributable to Cond. 20 performance, in which there were considerably more "last" than "other" responses recalled. This finding supports the "unlearning" results of Barnes and Underwood (1959). In addition, Cond. 5 and 10 also yielded more "last" than "other" responses recalled, although the difference was less than that of Cond. 20.

An analysis of variance performed upon the recognition-frequency data revealed that the RP, "last" vs. "other" response, and interaction sources of variation do not approach significance.

DISCUSSION

The results of the present experiment are considered in terms of probability learning, transfer, and the relation of these two areas.

The present experiment, within a probability-learning framework, is a two-event situation with $P(E_1) = P(E_2) = .5$. The independent variable was pattern of E_1-E_2 events with conditional E_1-E_2 probability ranging from $P(\text{repetition}) = 0.0$ in Cond. 1 to approximately .98 in Cond. 20. Since event and conditional-probability effects have been found previously in probability-learning experiments (e.g., Anderson, 1960; Zimmerman & Voss, 1962) the present study provided for extension of the range of conditional probability and patterns considered.

The results indicate that event-pattern variation had no significant effect on frequency of total anticipations of the two events. However, event pattern yielded a significant effect upon the relative R_1-R_2 anticipations. In particular, a primacy effect occurred, with the first presented response (R_1) yielding more anticipations than the second presented response (R_2) over Cond. R, 1, 5, and 10. In Cond. 20, R_2 exceeded R_1 frequency. Thus, with the .5-.5 response probability ratio of the present experiment, event-pattern variation yielded a significant effect upon the relative frequency of occurrence of the two responses. In addition, it is apparent from Fig. 1 data that the response sequences in Cond. 5, 10, and 20 generally followed the event patterns.

The present experimental design, within the A-B, A-C transfer framework, involves variation in A-B, A-C trial-block pattern. The results indicate that as A-B learning increases, transfer effects change from negative to positive. This finding is in agreement with other studies which show that negative transfer decreases as A-B learning increases (Jung, 1962; Mandler & Heinemann, 1956; Postman, 1962). A second transfer finding is that transfer effects diminish with repeated A-B, A-C shifts (McGeoch & Irion, 1952, pp. 338–339).

The recall and recognition findings are relevant to two aspects of the "unlearning" issue. First, the results indicate that "unlearning," defined as the lack of recall of B following A-B, A-C acquisition (Barnes & Underwood, 1959), is maximal in the traditional A-B, A-C design (Cond. 20). The difference of recall of "last" and "other" responses over Cond. 10, 5, and 1 further suggests that "unlearning" decreases with more frequent B and C response changes. Second, if the recognition data are assumed to indicate that no difference exists in A-B, A-C associative strength, then the results suggest that "unlearning" is due to a lack of the availability of the B response at the time of recall rather than to a loss of A-B associative strength. Furthermore, the more frequent shifts of B and C of Cond. 10, 5, and 1 provide for a greater availability of B and a correlated lack of "unlearning."

REFERENCES

ANDERSON, N. H. Effect of first-order conditional probability in a two-choice learning situation. *J. exp. Psychol.*, 1960, **59**, 73–93.

BARNES, J. M., & UNDERWOOD, B. J. "Fate" of first-list associations in transfer theory. *J. exp. Psychol.*, 1959, **58**, 97–105.

HORTON, D. L., & KJELDERGAARD, P. M. An experimental analysis of associative factors in mediated generalizations. *Psychol. Monogr.*, 1961, **75**(11, Whole No. 515).

JUNG, J. Transfer of training as a function of degree of first-list learning. *J. verbal Learn. verbal Behav.*, 1962, **1**, 197–199.

MANDLER, G., & HEINEMANN, S. H. Effect of overlearning of a verbal response on transfer of training. *J. exp. Psychol.*, 1956, **52**, 39–46.

MCGEOCH, J. A., & IRION, A. L. *The psychology of human learning.* New York: Longmans-Green, 1952.

MORNING, N., & VOSS, J. F. Mediation of R_1-R_2 in the acquisition of S-R_1, S-R_2 associations. *J. exp. Psychol.*, 1964, **67**, 67–71.

POSTMAN, L. Transfer of training as a function of experimental paradigm and degree of first-list learning. *J. verbal Learn. verbal Behav.*, 1962, **1**, 109–118.

VOSS, J. F., THOMPSON, C. P., & KEEGAN, J. H. Acquisition of probabilistic paired associates as a function of S-R_1, S-R_2 probability. *J. exp. Psychol.*, 1959, **58**, 390–399.

ZIMMERMAN, H. S., & VOSS, J. F. The effect of marginal and conditional probabilities of one verbal association upon concomitant acquisition of a second association. *J. verbal Learn. verbal Behav.*, 1962, **1**, 463–469.

(Received May 28, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 3, 309–314

INFORMATION AND REACTION TIME FOR "NAMING" RESPONSES [1]

ROBERT E. MORIN, ANDREW KONICK, NOLA TROXELL, AND SANDRA McPHERSON

Kent State University

The relationship between stimulus information (1, 2, or 3 bits) and reaction time (RT) was investigated for 5 types of associations all of which required naming responses to familiar stimuli (Experiment I—faces of fraternity brothers; Experiment II—line drawings of animals; Experiment III—colors; Experiment IV—geometric symbols; Experiment V—letters of the alphabet). Each of 54 Ss served in 3 information conditions of 1 of the 5 experiments. The observed slope for letters, though reliable, was less than .01 sec/bit. In Experiments I–IV marked increases in RT accompanied changes in stimulus information, and all curves were negatively accelerated. The results indicate that a high degree of overlearning is not sufficient to produce independence of RT and information measures. A response-competition hypothesis is suggested to account for differences between the results with letters and other stimuli.

For many years the empirical generalization that choice reaction time (RT) increases with the size of the stimulus set was among the most firmly established psychological principles. From the early work of Merkel (1885) to recent research relating information measures to RT, support for the generalization has been found with a wide variety of tasks (Brainard, Irby, Fitts, & Alluisi, 1962; Bricker, 1955; Hick, 1952; Hyman, 1953). In terms of information measures, these results have been summarized by the proposition that RT is a positive linear function of the average information transmitted per response (H_t).

Within the past 5 yr. several studies have shown that there are some tasks for which RT apparently does *not* increase as a function of stimulus or transmitted information. Leonard (1959) observed this result when *S*s made finger-movement responses to tactual stimuli applied directly to the finger tips. David, Moray, and Treisman (1961) required *S*s to repeat or "imitate" tape-recorded sounds played through earphones. The variable of number of stimulus alternatives did not influence RT when the stimuli were numbers, letters, or words. Mowbray (1960) found that the time to name visually presented numerals remained a constant over the range of 2–10 stimuli. Similar results have been reported by Brainard et al. (1962) and Morin and Forrin (1962). In another study, Mowbray and Rhoades (1959) used a single *S* and observed that initial differences between finger reactions to two and four lights disappeared after 39,000 trials of practice.

In most of the studies cited the investigators have expressed the view that practice, familiarity, and/or stimulus-response compatibility are important variables which influence whether or not the RT-H_t function will have zero slope. Though this conclusion seems appropriate to the tasks and results upon which it has been based, the generality of the posi-

---

[1] This study was supported by National Science Foundation Grant GB-323 awarded to Kent State University.

TABLE 1

RESPONSE LISTS FOR FIVE EXPERIMENTS

I Faces	II Animals	III Colors	IV Symbols	V Letters
BILL	BIRD	GREEN	SQUARE	F
DON	FROG	PINK	DOTS	G
JOE	PIG	BLUE	CIRCLE	A
AL	GOAT	YELLOW	TRIANGLE	E
DAN	DOG	VIOLET	CROSS	B
JACK	HORSE	RED	LINE	C
TOM	CAT	ORANGE		D
RON	SEAL	WHITE		H

tion would be more firmly established were it founded on a broader sample of tasks. All adults have available a large repertoire of highly overlearned responses; there are many objects and symbols which they can easily name. The purpose of the present investigation was to determine whether independence of RT and information measures is maintained over a variety of associations all of which require "naming" responses to familiar stimuli.

METHOD

Five experiments were conducted and in each a different class of stimuli was used. These were: Exp. I—faces of fraternity brothers; Exp. II—line drawings of animals; Exp. III—colors; Exp. IV—geometric symbols; Exp. V—letters of the alphabet. Though the studies were similar enough to be regarded as conditions of a single investigation, they are referred to as separate experiments primarily because they were not run at the same time by the same *E*. In addition, there were minor variations in procedure and design.

Subjects.—From a total of 54 undergraduate students at Kent State University, 13, 12, 8, 10, and 11 served as *S*s in Exp. I–V, respectively. Males were used in Exp. I, II, and V, and females in Exp. III. In Exp. IV there were six males and four females. For their participation *S*s either were credited with fulfilling an introductory psychology course requirement or were paid $1.00.

Apparatus.—Stimuli in Exp. I–III were presented by a Kodak Carousel slide projector positioned to the rear of a milk-glass viewing screen located 2 ft. from *S*. Projected images were approximately 6.5 × 10.0 in. in Exp. I and III, and 4.0 × 4.0 in. in Exp. II. In the final two experiments symbols and letters varying between .5 and 1.0 in. in their principal dimensions appeared on Industrial Electronic Engineers display cells (Series 10,000). The cells were mounted 1 ft. from *S* in a black "shadow box" and were programed by an eight-channel tape reader. A Hunter voice key, a Hewlett-Packard electronic counter, and a Hewlett-Packard digital recorder measured and recorded RT.

Stimulus materials.—Table 1 lists the responses associated with the set of stimuli employed in each experiment. For Exp. I slides were prepared from black and white portrait-style photographs of fraternity brothers. Since *S*s were all actives living in the same fraternity house, each man pictured was well-known by his first name. The eight men whose names appear in Table 1 also served as *S*s. When one did so, slides depicting him were replaced by those of another brother named JIM. In Exp. II the stimuli were white line drawings of animals on a black ground. The colors used in Exp. III were obtained from cellophane placed in slide binders and projected on the milk-glass screen. Some properties of the stimuli in Exp. IV are not apparent from the names given to them: the TRIANGLE was equilateral and rested on its base; the three DOTS were equally spaced and horizontally arranged; the vertical and horizontal legs of the CROSS were equal in length and met at their midpoints; the LINE was horizontal.

In each experiment the intent was to select readily identifiable visual stimuli for which *S*s had highly overlearned single word responses. Thus, for most *S*s the word TRIANGLE is clearly a dominant response to a visually presented triangle, though certainly other responses exist at lower associative strengths (e.g., pyramid, rack).

Experimental conditions.—The three conditions in each experiment were distinguished in terms of number of stimulus alternatives or average information per stimulus (H_s). With a single exception there were 2, 4, or 8 stimuli in a condition, producing 1, 2, or 3 bits of information per stimulus, respectively. In Exp. IV, six geometric symbols (2.58 bits per stimulus) were used in lieu of a condition with eight symbols.

The responses listed in Table 1 may be interpreted to indicate which stimuli were employed in each condition. When n stimuli (2, 4, 6, or 8) were needed in a condition, the stimuli associated with the first n responses of the appropriate list in Table 1 were always the ones chosen. Thus, for the condition in Exp. II with four stimuli (2 bits of stimulus information), slides depicting a bird, a frog, a pig, and a goat were used for all *S*s.

Procedure.—At the beginning of the single experimental session *S* was shown and asked to name, by a single word, each stimulus in the total set for the experiment. Correspondence between the names provided by *S*s and those listed in Table 1 was perfect for stimuli in conditions with 1 or 2 bits of information. When discrepancies did arise with other stimuli (e.g., the cross was called PLUS and violet was called PURPLE), *S* used his own responses and not those in Table 1.

Each *S* served in all three conditions. The experimental session was divided into nine blocks of practice. Within successive sets of three blocks, each condition occurred once, but no condition ever followed itself. Apart from these restrictions the order of conditions was randomly determined for each *S*.

A block of trials consisted of 72 stimulus presentations. The eight initial stimuli were randomly varied "lead-ins" used for warm-up and to disguise the fact the terminal 64 stimuli were always in the same order for a given condition. Sequences for the final 64 stimuli equated the frequency of each stimulus within a condition and first-order transitional probabilities as well. Furthermore, parallels existed between the stimulus orders used in the three conditions. In each study there were two critical stimuli, the two common to all conditions. If a critical stimulus was presented on Trial *k* in a condition with eight stimuli, the same stimulus was presented on Trial *k* in the other two conditions.

When a stimulus appeared it remained on until *S*'s response activated the voice key. In the experiments employing the slide projector the interval between the termination of one stimulus and the onset of the subsequent stimulus was .7 sec. When display cells were used, the comparable period was 1.0 sec. During the 1.0-min. rests between blocks of trials, *E* changed slide trays or stimulus tapes, and reminded *S* to react as rapidly as possible consistent with a high degree of accuracy.

RESULTS

Since there were no changes in the relationships among conditions with stage of practice, all RT data reported are based on a single measure for each *S* in each condition. This measure was a mean for 48 reactions to critical

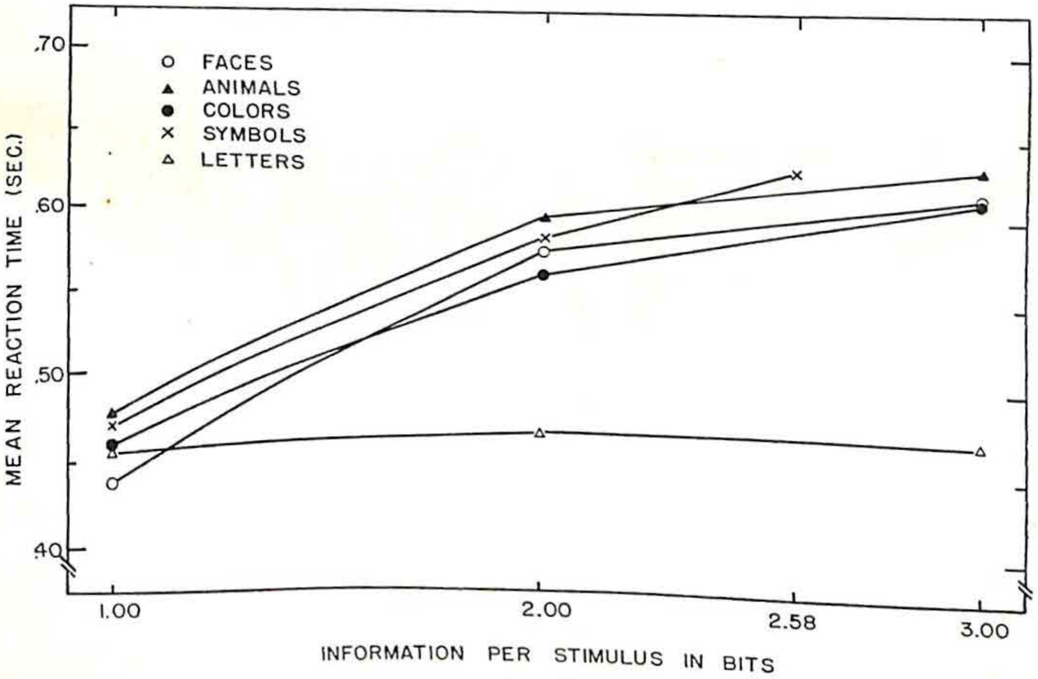


FIG. 1. Mean RT in seconds as a function of information per stimulus in bits for the five experiments.

TABLE 2

Summary of Trend Analyses for Five Experiments

Exp.	Source	*df*	*MS*	*F*
I Faces	Slope	1	1980.14	673.52**
	Quadratic	1	245.93	73.19**
	Ind. slope	12	2.94	
	Ind. quad.	12	3.36	
II Animals	Slope	1	1409.13	98.96**
	Quadratic	1	169.59	33.06**
	Ind. slope	11	14.24	
	Ind. quad.	11	5.13	
III Colors	Slope	1	950.18	78.92**
	Quadratic	1	43.13	21.78**
	Ind. slope	7	12.04	
	Ind. quad.	7	1.98	
IV Symbols	Slope	1	1348.96	88.40**
	Quadratic	1	14.24	9.56*
	Ind. slope	9	15.26	
	Ind. quad.	9	1.49	
V Letters	Slope	1	11.93	6.28*
	Quadratic	1	2.72	1.85
	Ind. slope	10	1.90	
	Ind. quad.	10	1.47	

* $p < .05$.
** $p < .01$.

stimuli, 16 from each of three blocks of practice allocated to a given condition. The 16 trials selected were the ones for which critical stimuli were the same in all three conditions.

In Fig. 1 mean RT is plotted as a function of H_s for the three conditions of each of the five experiments. Most notable are the marked increases in RT with H_s for faces, animals, colors, and symbols. Considerably smaller increases averaging less than .01 sec/bit were observed for letters. Trend tests analyzing the significance of the H_s variable in each experiment are summarized in Table 2. In all studies the linear component of the regression of RT on H_s was significant when evaluated against the variability of individual slopes. The finding of statistical significance for letters reflects the homogeneity of different *S*s on the slope measure. The quadratic components of the obtained functions were significant in Exp. I–IV, a result predictable from the finding that the curve for every *S* in these experiments was negatively accelerated.

Error rates were relatively low in all experiments even though any incorrect verbal response which set off the voice key was defined as an error. Thus, some errors were inadvertant "Ah's," and others were hybrid responses. Table 3 presents percentage of error for reactions on critical trials in each condition of the five studies. The maximum observed error rate of 2.3% represents an average of slightly more than one error per *S* for the 48 critical trials within a given condition. For no *S* were errors sufficiently frequent to produce more than minor discrepancies between H_t and H_s.

Discussion

The most striking finding in the present series of studies was the failure, in Exp. I–IV, to find functions with zero or near-zero slope when RT was plotted against H_s. Only when the stimuli were letters did the obtained slope approach zero. If the associations in the first four experiments may properly be described as highly overlearned, then it is apparent that overlearning is not a sufficient condition to produce independence of RT and H_t or H_s.

One obvious question concerns the distinguishing characteristics of associa-

TABLE 3

Mean Percentage of Error on Critical Trials

H_s	Exp.				
	I Faces	II Animals	III Colors	IV Symbols	V Letters
1	1.9	0.2	0.0	0.8	0.8
2	1.6	2.1	1.0	2.3	0.6
3	1.0	1.2	0.8	2.1	0.0

tions producing each type of result. What is different about naming letters (or numerals) from naming the faces of friends, pictures of animals, colors, or geometric symbols? Among the potentially important distinguishing variables are stimulus discriminability, degree of overlearning, and amount of response competition. Of the three, response competition tentatively appears most relevant.

Letters may be more discriminable than faces or colors, but the same conclusion would not seem to apply to a comparison of letters and geometric symbols. The names of letters probably are more highly overlearned than any of the other responses, and this could be a reason for the observed differences in performance. On the other hand, adults who started naming animals and colors at a preschool age should know these names at least as well as first-grade children know the names of numerals. Yet Morin and Forrin (1965), in a study with numerals as stimuli, found independence of RT and H_s both for first and third graders.

A hypothesis related to, but distinct from, degree of overlearning is that classes of associations differed in terms of the number and strength of responses competing with each dominant response. Despite a very high associative strength between the picture of a "dog" and the response DOG, *S* might find it somewhat easier to rename this stimulus (e.g., COLLIE, POOCH, LASSIE) than to attach a second, third, or fourth name to a letter of the alphabet or a numeral. The word CIRCLE was the first response to the "circle," but it would not have been incorrect to have called the stimulus BALL, DISC, or COIN. Response competition was also possible for faces and colors. Some of the men had nicknames, and the dominant response to the colors may conceivably have competed with words which described subtler aspects of hue and saturation for the female *S*s.

Three other features of the data are noteworthy. First, RTs for the groups with 1 bit of stimulus information were remarkably similar considering differences in *S*s, procedure, and initial sounds of critical stimuli. The observed differences in RT were not significant; $F(4, 49) = .79$. With only two alternative responses, and these highly overlearned, perhaps *S* can raise both to a near threshold level prior to stimulus presentation and thus avoid the effects of response competition.

A second point of interest is the observation of a small but significant slope for letters where independence might have been anticipated. This finding from Exp. V is not inconsistent with the view of Fitts and Switzer (1962) who would say that a positive slope should be observed, even for letters or numerals, provided the smaller subsets of stimuli are relatively familiar.

A final point of significance was the finding of nonlinear functions in Exp. I–IV. Linearity is the more common result. On the average, in these studies, increasing H_s from 1 to 2 bits added .121 sec. to RT, but a change from 2 to 3 bits cost only .065 sec. Both of these values are low relative to those obtained in other research (Brainard et al., 1962) with verbal reactions (excluding numeral naming and letter naming). The two findings, nonlinearity plus a somewhat smaller slope, may both be associated with the high degree of overlearning which characterized all four tasks. Seibel (1959) gave many thousands of trials of practice on key-pressing responses to light stimuli and concluded from terminal performance that "RT is an increasing but negatively accelerated function of the log of the number of alternatives, with little or no increase beyond approximately eight alternatives." One other possibility is that curvilinearity might be associated with the use of *S*s as their own controls and/or with differential practice on stimuli common and stimuli not common to all conditions. Several analyses, including a *between-Ss* comparison made prior to differential practice, indicate that the nonlinear functions were not simply a consequence of the design and procedure employed.

Independently of the present research, Fraisse (1964) has conducted very similar investigations. Working with colors and letters, his curves are practically identical to those in Fig. 1. It is likewise his interpretation that overlearning is not a sufficient condition to produce independence of RT and number of stimulus alternatives.

REFERENCES

Brainard, R. W., Irby, T. S., Fitts, P. M., & Alluisi, E. A. Some variables influencing the rate of gain of information. *J. exp. Psychol.*, 1962, **63**, 105–110.

Bricker, P. D. Information measurement and reaction time. In H. Quastler (Ed.), *Information theory in psychology.* Glencoe, Ill.: Free Press, 1955. Pp. 350–359.

Davis, R., Moray, N., & Treisman, A. Imitative responses and the rate of gain of information. *Quart. J. exp. Psychol.*, 1961, **13**, 78–89.

Fitts, P. M., & Switzer, G. Cognitive aspects of information processing: I. The familiarity of S-R sets and subsets. *J. exp. Psychol.*, 1962, **63**, 321–329.

Fraisse, P. Relations entre la durée du seuil de reconnaissance perceptive et le temps de réaction verbale. [Relations between the threshold of perceptive recognition and the time of verbal reaction.] *Proceedings, XVIIth International Congress of Psychology.* Amsterdam: North-Holland, 1964. Pp. 197–198.

Hick, W. E. On the rate of gain of information. *Quart. J. exp. Psychol.*, 1952, **4**, 11–26.

Hyman, R. Stimulus information as a determinant of reaction time. *J. exp. Psychol.*, 1953, **45**, 188–196.

Leonard, J. A. Tactual choice reactions: I. *Quart. J. exp. Psychol.*, 1959, **11**, 76–83.

Merkel, J. Die zeitlichen Verhaltnisse der Willensthatigkeit. *Phil. Stud.*, 1885, **2**, 73–127.

Morin, R. E., & Forrin, B. Mixing of two types of S-R associations in a choice reaction time task. *J. exp. Psychol.*, 1962, **64**, 137–141.

Morin, R. E., & Forrin, B. Information processing: Choice reaction times of first- and third-grade students for two types of associations. *Child Develpm.*, 1965, in press.

Mowbray, G. H. Choice reaction times for skilled responses. *Quart. J. exp. Psychol.*, 1960, **12**, 193–202.

Mowbray, G. H., & Rhoades, M. V. On the reduction of choice reaction times with practice. *Quart. J. exp. Psychol.*, 1959, **11**, 16–23.

Seibel, R. Discrimination reaction time as a function of the number of alternatives and of the particular stimulus-response patterns. *Amer. Psychologist*, 1959, **14**, 396. (Abstract)

(Received June 1, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 3, 315–322

EXTENDED TRAINING EXTINCTION EFFECT UNDER MASSED AND SPACED EXTINCTION TRIALS[1]

DAVID BIRCH

University of Michigan

The previously reported negative relationship between extent of acquisition and response speed during extinction on the straight runway when large numbers of training trials are used was investigated under markedly different extinction intertrial intervals. 2 degrees of training, 60 or 190 trials, were factorially combined with 2 extinction intertrial intervals, 1 min. or 24 hr., and extinction was evaluated over 3 5-trial blocks separated by at least 48 hr. Results for the postrecovery speed measure showed that 190 acquisition trials led to faster extinction than 60 trials for both extinction intertrial intervals. Also, under massed extinction decrements within a 5-trial block decrease during extinction following 60 acquisition trials but increase following 190 trials.

Experiments by North and Stimmel (1960) and Birch, Ison, and Sperling (1960) produced data which indicate that the extinction of the approach response in the straight runway is not a monotonically increasing function of the number of acquisition trials. In both cases during extinction the starting speed decreased more rapidly and reached a lower level for the groups given greater numbers of acquisition trials. Later, Ison (1962) and Clifford (1964) systematically varied the number of training trials across groups and found a nonmonotonic relation between mean speed of running during extinction and extent of acquisition. The form of this function is an initially rising and subsequently falling one (Birch, 1961). Ison trained his rats for either 10, 20, 40, 60, 80, or 100 trials and found fastest response in extinction for his 20-trial group. Clifford gave 24, 48, 84, 132, or 192 acquisition trials with the 48- and 84-trial groups showing fastest response in extinction.

Hill and Spear (1963) failed to find nonmonotonicity with groups given 8, 16, 32, 64, or 128 acquisition trials, and House (1963) reports monotonicity for acquisition groups of 30, 60, and 120 trials. Theios and Brelsford (1964), however, specifically attempted to reproduce the conditions of Ison's experiment and replicated, almost to the point of identical values, the findings for his 10- and 60-trial groups.

Experimentation so far has not isolated and identified the conditions of acquisition and extinction under which nonmonotonicity holds. One condition of possible importance is the intertrial interval imposed during extinction. In four of the five experiments showing the extended training extinction effect (Clifford, 1964, is the exception) a small number of relatively spaced extinction trials were given within a day, and extinction was continued over a number of days. In the two experiments not showing the effect, extinction trials were given under more massed conditions in only 1 or 2 days. Thus, there is an indication that the extended training extinction effect is more likely to be observed under a series of spaced

---

[1] This research was supported by National Science Foundation Grant G 19217 to the author.

extinction days than it is under a single day of massed extinction.

The present experiment was designed to pursue this point by factorially combining two levels of training, 60 and 190 trials, chosen on the basis of Clifford's (1964) results, with two widely different extinction intertrial intervals, 24 hr. and approximately 1 min. In addition, a special arrangement of grouping extinction trials into sets of five with a prolonged rest between sets was devised. This blocking of the extinction trials allows comparison of the training groups in terms of the temporary extinction occurring within sets of five trials and the permanent extinction evaluated after opportunity for spontaneous recovery.

METHOD

Subjects.—The *S*s were 57 naive male rats, approximately 90 days old at the beginning of training, from the colony maintained by the Michigan Psychology Department. Two *E*s,[2] each of whom collected half the data in each condition, began with 28 and 29 *S*s. A total of 8 *S*s was discarded for excessively slow running and failure to eat and 1 *S* died. All losses occurred during acquisition with the distribution 2, 4, 1, 1, 1 on Days 1, 2, 3, 4, and 10. The experiment was completed with 6 *S*s in each of four conditions for each *E*.

Apparatus.—The straight alley described by Clifford (1964) was used. Briefly, the alley is enclosed, 52½ in. long, 4 in. wide, and 6 in. high, covered with ¼-in. hardware cloth and is painted flat black. A laterally sliding door, 11 in. from the front of the alley defines the start compartment and a shaped aluminum food dispenser, which prevented sight of the food cup until *S* was on top of it, was attached to the end of the alley. No goal-box door was employed. Four clocks, timing *S* in the start box and over three additional 10-in. sections of the alley, were controlled by a system using infrared photocell beams.

Procedure.—The *S*s were deprived of food over a weekend and then maintained on a 22-hr. schedule with 10–12 gm. of laboratory chow for the duration of the experiment. Pretraining took 5 days during which each *S* was handled each day, allowed to drink 10 drops of 20% sucrose solution on each of the last 4 days, permitted to explore the alley for 10 min. on the fourth day, and given one preliminary acquisition trial on the fifth day.

Acquisition training was given only on weekdays with *S*s fed on their regular maintenance schedule on Saturday and Sunday. Five trials per day were run on the first 2 training days and 10 per day thereafter. On each trial *S* was confined in the start box for 15 sec. before the start-box door was opened and was left in the alley for 30 sec. following his consumption of the two drops of 20% sucrose solution. The start-box door was closed while *S* was drinking. Water was available in the home cages at all times and also in the holding cages between trials. The minimal intertrial interval was about 1 min.

Extinction occurred with the food cup in place but empty. In the massed condition *S* was given five trials in immediate succession with an intertrial interval of about 1 min. The *S* was returned to the holding cage between trials while *E* read the clocks. Spaced extinction was administered at one trial a day. Acquisition ended on Friday and extinction began on Monday for the spaced groups. A block of five trials carried through Friday, and no trials were run on the weekend. Three such blocks of five spaced trials were provided plus a sixteenth trial the final Monday. This procedure defined blocks of five spaced trials separated by 72 hr. For the massed groups acquisition ended on a Tuesday, and extinction began the following Friday giving the same 2-day interval as for the spaced groups. The *S* received a set of five massed trials on Friday, second and third sets on the following Monday and Wednesday, and a sixteenth trial on Friday. This procedure defined blocks of five massed trials separated by 72 or 48 hr. The day of the week on which acquisition began was chosen to result in the desired starting days for extinction.

Design.—Two degrees of acquisition, 60 or 190 trials, are combined factorially with two conditions of extinction, massed or spaced trials. Each of the four cells contains 12 *S*s, 6 of which were run by each of two *E*s.

RESULTS

During alley training the groups differed only in the number of trials, either 60 or 190, received. The acquisition data for all four groups

---

[2] James Allison and L. Thomas Clifford.

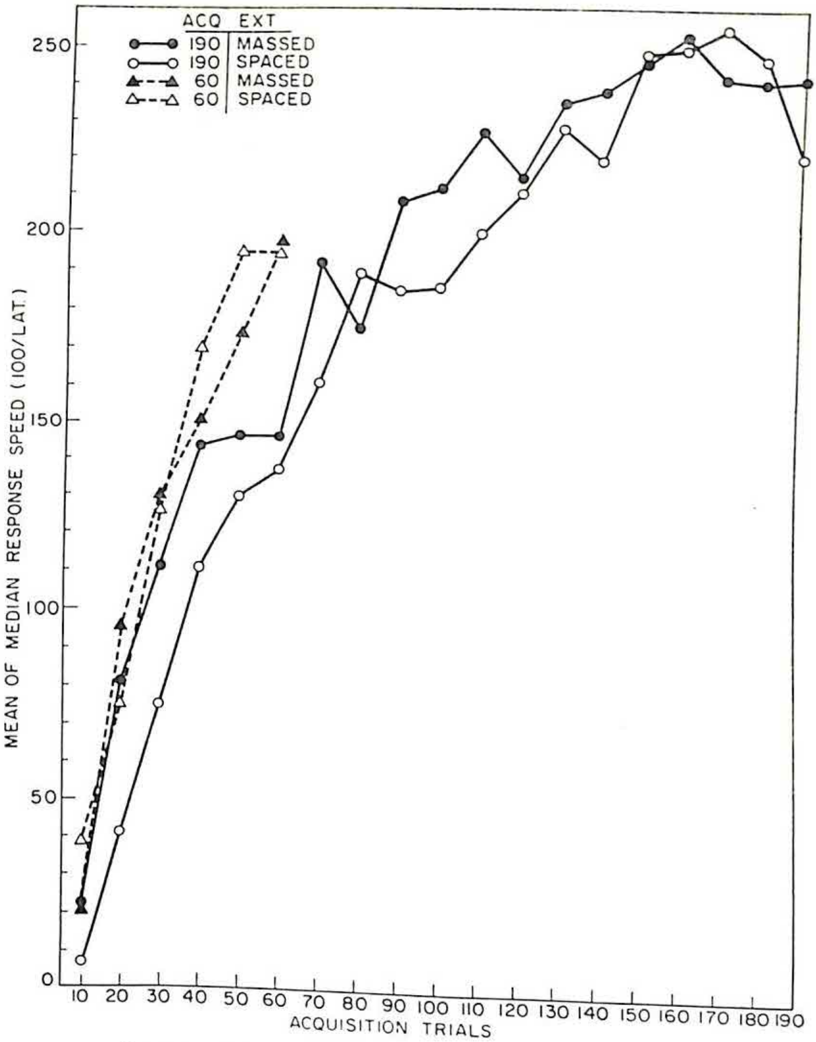


FIG. 1. Mean response speed during acquisition.

pertinent to extinction are shown in Fig. 1. Each point is the mean response speed of leaving the start box for the 12 *S*s of a group where each *S*'s score is the median of the last 9 trials of the day. Groups 60M and 60S are similar to each other as are Groups 190M and 190S, but there is a suggestion of relatively faster acquisition in the two 60-trial groups than in the 190-trial groups.

Terminal acquisition was evaluated by an analysis of variance on the data from the last training day. Amount of training, future extinction condition, and *E* form a three-dimensional factorial design. The analysis shows a significant main effect for Amount of Training, $F\ (1, 40) = 5.89$, $p < .05$. The other main effects and the interactions yield $F < 1$. Thus, the four basic groups enter extinction differing only in terminal acquisition speed attributable to extent of training. When the data from Trial 1 of extinction are used to evaluate terminal acquisition response-speed differences, however, none of the conditions, either singly or in interactions, show significance.

Mean response speed over the 16 extinction trials is shown in Fig. 2. The most striking aspects of the data are (*a*) the relatively steeper drop in the postrecovery speeds (Trials 1, 6, 11, and 16) for the massed and spaced 190-trial groups as compared with the corresponding 60-trial groups, (*b*) the lack of differences in postrecovery speeds for the massed and spaced extinction conditions, and (*c*) the marked effect of the extinction intertrial interval on response speed within a five-trial block. Table 1 contains the mean postrecovery response speeds (Trials 1, 6, 11, and 16) and the mean response speed decrements within the five-trial blocks (Trials 1–5, 6–10, and 11–15) for the four treatment groups.

The effects of extent of training, extinction intertrial interval, and *E* on postrecovery response speed were evaluated from Trials 1, 6, 11, and 16 using a four-way analysis of variance with trials as a correlated dimension. The pattern of results shows that the postrecovery response speeds are affected by the extent of training but not by the intertrial interval in extinction. The interaction between

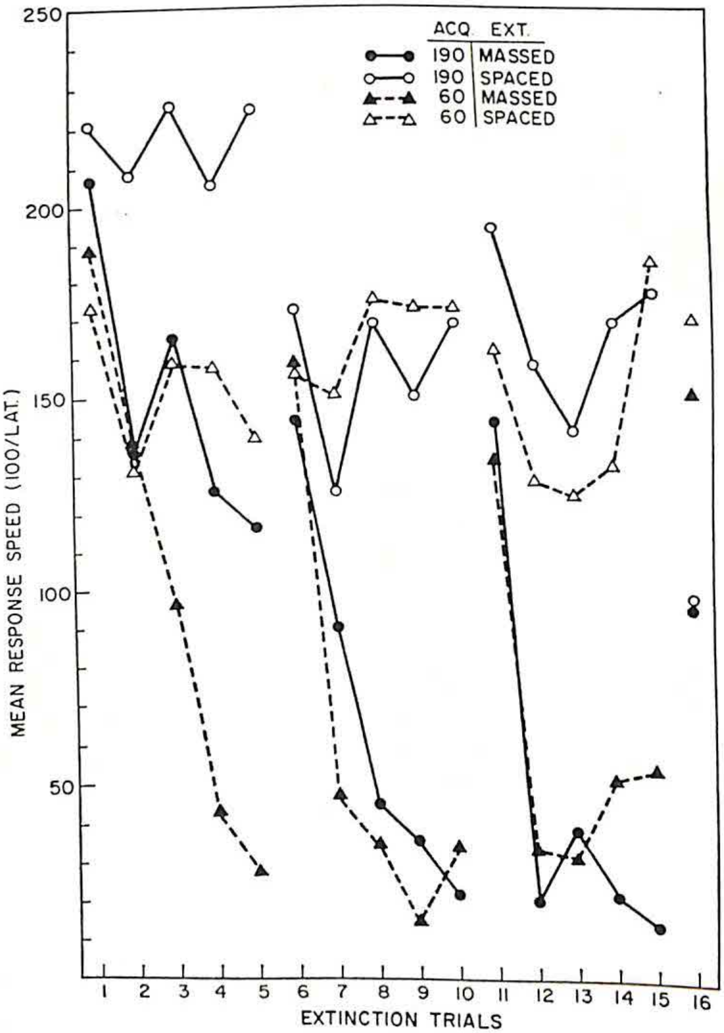


FIG. 2. Mean response speed in extinction as a function of extent of training and extinction intertrial interval.

TABLE 1

MEAN POSTRECOVERY RESPONSE SPEED (TRIALS 1, 6, 11, AND 16) AND MEAN RESPONSE-SPEED DECREMENTS (TRIALS 1–5, 6–10, AND 11–15)

Trials	Postrecovery Response Speed				Decrement		
	1	6	11	16	1–5	6–10	11–15
190M	207	143	143	96	90	121	129
60M	188	157	133	150	159	124	78
190S	221	172	193	98	−3	4	17
60S	173	155	161	169	34	−17	−23

Extent of Training and Extinction Trials was significant, F (3, 120) = 5.07, $p < .01$, indicating the course of extinction was different following 60 and 190 acquisition trials. In addition, *S*s given 190 acquisition trials were found to be significantly slower than those given 60 trials, F (1, 44) = 7.77, $p < .01$, when evaluated on Trial 16 alone.

Neither extent of training nor extinction intertrial interval as main effects nor their interaction approached significance. In fact, except for the Trials main effect, F (3, 120) = 8.90, $p < .001$, the only other significant sources of variance were the interactions between Trials and *E*, F (3, 120) = 2.91, $p < .05$, and among Extent of Training, Trials, and *E*, F (3, 120) = 5.51, $p < .01$. All other F values tended to be close to or less than unity. Examination of the data shows that the Trials × *E* interaction arises because the extinction is occurring at different rates for the two *E*s. Similarly, the triple interaction is present because of differences in degree of the first-order interaction between Extent of Training and Extinction Trials for the two *E*s rather than because of differences in the form of the interaction.

To facilitate evaluation of the response-speed decrements within blocks of five trials, mean values for differences in speed between Trials 1 and 5, 6 and 10, and 11 and 15 have been entered in Table 1. The data on decrements were analyzed by a four-way analysis of variance with trial blocks as a correlated dimension.

Five massed trials clearly result in larger decrements than five spaced trials, F (1, 40) = 52.23, $p < .001$. In addition, the trend in the size of decrements over trial blocks is different for the two training conditions as shown by the significant Extent of Training × Trial Blocks interaction, F (2, 80) = 4.17, $p < .05$. Examination of Table 1 reveals that for Groups 190M and 190S decrements get larger as extinction proceeds while for Groups 60M and 60S decrements get smaller. The only other significant source of variance in the analysis is attributable to the main effect for *E*s, F (1, 40) = 6.94, $p < .05$.

Results for running speed parallel those for response speed. Readings from the last three clocks were summed to give *S*'s total running times which were converted to running speeds by a reciprocal transformation. Terminal acquisition running speed, when evaluated by analysis of variance on the last training day, showed all main effects, Extent of Training, Extinction Intertrial Interval, and *E*, nonsignificant. The only significant effect was for the interaction between Extent of Training and Extinction Intertrial Interval. This effect occurs

TABLE 2

Mean Postrecovery Running Speed (Trials 1, 6, 11, and 16) and Mean Running-Speed Decrements (Trials 1–5, 6–10, and 11–15)

Trials	Postrecovery Running Speed				Decrement		
	1	6	11	16	1–5	6–10	11–15
190M	89	59	78	42	30	29	46
60M	118	88	92	91	89	49	52
190S	104	54	53	33	27	−12	−15
60S	108	101	107	77	16	−26	0

primarily because the 190S group ran unusually slowly. A comparable three-factor analysis of variance for Trial 1 of extinction produced nonsignificance throughout.

Table 2 contains the postrecovery and decrement data for running speed. These data, like those for response speed, show a greater postrecovery extinction effect following 190 acquisition trials than following 60, little difference in postrecovery speeds for massed and spaced extinction, and marked effects of extinction intertrial interval on decrements within a 5-trial block. It might also be noted from Table 2 that the mean postrecovery running speeds for the 190-trial groups are appreciably below those for the 60-trial groups by Trial 6 after being only slightly below on Trial 1. This contrasts with the response speed results which indicate it takes until Trial 16 for the separation between the training groups to occur.

Analysis of variance on the postrecovery running-speed data shows a pattern of significance similar to that for response speed with one exception. For running speed the Extent of Training main effect was found to be significant, $F\ (1,\ 40) = 20.89$, $p < .001$, and the interaction between Extent of Training and Extinction Trials nonsignificant while for response speed the opposite held. That the effect attributable to training appeared early in extinction for the running-speed measure is illustrated by the nonsignificant difference between the 190- and 60-trial groups on Trial 1, and the significant difference between these groups on Trial 6, $F\ (1,\ 40) = 16.71$, $p < .001$.

Analysis of variance was also carried out on the running-speed decrements presented in Table 2. In correspondence with the response-speed results, differences between the massed and spaced extinction intertrial intervals were highly significant, but no significant effects attributable to Extent of Training were found. Running-speed decrements tend to become smaller for both training groups as extinction proceeds yielding a significant Trial Blocks main effect, $F\ (2,\ 80) = 7.57$, $p < .01$, and a nonsignificant Extent of Training × Trial Blocks interaction. These results are opposite in pattern to those for response speed. Significant effects due to E were found at both the main effect and interaction levels but examination showed these effects to be in terms of magnitude rather than direction.

Discussion

The data agree with previous straight-alley findings of a negative relationship between the extent of acquisition and approach response speed during extinction when large numbers of training trials are used. Groups given 190 acquisition trials showed a faster rate of extinction than those given only 60 trials for both postrecovery speed measures. These postrecovery, or what might appropriately be called permanent, extinction effects were not affected by the massing or spacing of the five trials within a block, however. Thus, it appears that whatever may be the source of the extended training extinction effect, it operates on permanent extinction and does so independently of the intertrial interval used during extinction.

Differences as a function of extent of training also show up in the response-speed decrements that occur within a five-trial block under massed extinction. These decrements decrease in magnitude across blocks for Group 60M but increase for Group 190M. A similar though less clear pattern holds for running speed.

The results for both the postrecovery and the decrement measures are consistent with a two-factor theory of the extinction of the instrumental approach response in the runway offered by Birch (1961). This theory grows out of those of Amsel (1958) and Spence (1960) and has much in common with Amsel's (1962) recent statement. It is proposed that the approach response in the alley declines both because of the experimental extinction of the fractional anticipatory goal response (r_g-s_g) supporting the instrumental response and because of the introduction of the antagonistic effects of frustration (r_f-s_f). The magnitude of the frustration effect is assumed to be an increasing function of the strength of r_g-s_g present when nonreinforcement occurs.

The more abrupt decline in response speed and in running speed for groups given the 190 training trials is said to occur because r_g-s_g is stronger for these groups than for groups given only 60 training trials and as a result proportionally more of the total extinction is attributable to the antagonistic effects of frustration. This same argument can be extended to account for the decrements in response speed occurring within the five trials of massed extinction for Groups 190M and 60M. As extinction of the instrumental response proceeds, the decrements become progressively smaller for Group 60M because the effect is due primarily to the negatively accelerated course of experimental extinction of r_g-s_g. In Group 190M, however, the additional action of r_f-s_f leads to increasing decrements over the number of extinction trials used. This is because the frustration response, elicited in the goal chamber, must occur in the start box in order to be effective in reducing the start speed. Presumably, a number of trials are required for the conditioning of r_f-s_f to the start-box cues, and response-speed decrements become larger as this conditioning proceeds. This trend in size of decrements would be expected only in the initial stages of extinction when r_g-s_g and r_f-s_f are both strong. The similar, though less pronounced, difference in trend between 190M and 60M for the running-speed measure can be accounted for in the same way.

Since the 60S and 190S groups show the same kind of opposite trends in response-speed decrements as are shown by the 60M and 190M groups, it appears that the effects attributed to frustration may not depend on the massing of the extinction trials. Apparently the magnitude of the frustration reaction and the course of conditioning for the effects of this reaction to cues of the start box are not affected by the intertrial intervals used. The result of massing extinction trials is to increase the size of the decrements within a 5-trial block over the spaced trials condition but to do so relatively constantly across blocks of extinction trials and equally for the two acquisition conditions. This result tends to hold for running speed as well as for response speed though there is a suggestion of differences in trend for the two 190-trial groups.

Two further facts in the data are noteworthy. First, the response speeds for Group 60S actually increase over the five trials within the second and third blocks of extinction rather than simply decrease less, as is true for Group 60M. Secondly, Groups 60M and 60S show greater initial response-speed decrements (Trials 1–5) than their counterparts, 190M and 190S. Both of these results relate to finer aspects of the data than have previously been considered by the theory and raise questions concerning the details of the process going on early in extinction.

REFERENCES

AMSEL, A. The role of frustrative nonreward in noncontinuous reward situations. *Psychol. Bull.*, 1958, **55**, 102–119.

AMSEL, A. Frustrative nonreward in partial reinforcement and discrimination learning: Some recent history and a theoretical extension. *Psychol. Rev.*, 1962, **69**, 306–328.

BIRCH, D. A motivational interpretation of extinction. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1961.* Lincoln: Univer. Nebraska Press, 1961.

BIRCH, D., ISON, J. R., & SPERLING, S. E. Reversal learning under single stimulus presentation. *J. exp. Psychol.*, 1960, **60**, 36–40.

CLIFFORD, T. Extinction following continuous reward and latent extinction. *J. exp. Psychol.*, 1964, **68**, 456–465.

HILL, W. F., & SPEAR, N. E. Extinction in a runway as a function of acquisition level and reinforcement percentage. *J. exp. Psychol.*, 1963, **65**, 495–500.

HOUSE, R. F. *Resistance to extinction as a function of sucrose concentration and number of acquisition trials.* (Doctoral dissertation, University of Michigan) Ann Arbor, Mich.: University Microfilms, 1963, No. 64-6692.

ISON, J. R. Experimental extinction as a function of number of reinforcements. *J. exp. Psychol.*, 1962, **64**, 314–317.

NORTH, A. S., & STIMMEL, D. T. Extinction of an instrumental response following a large number of reinforcements. *Psychol. Rep.*, 1960, **6**, 227–234.

SPENCE, K. W. *Behavior theory and learning.* Englewood Cliffs, N. J.: Prentice-Hall, 1960.

THEIOS, J., & BRELSFORD, J. Overlearning-extinction effect as an incentive phenomenon. *J. exp. Psychol.*, 1964, **67**, 463–467.

(Received June 2, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 3, 323–327

SELECTION STRATEGIES IN CONCEPT ATTAINMENT AS A FUNCTION OF NUMBER OF PERSONS AND STIMULUS DISPLAY[1]

PATRICK R. LAUGHLIN

Northwestern University

The selection strategies of individuals and 2-person cooperative groups were investigated in 5 concept-attainment problems. 2 types of stimulus displays were used: (a) form displays, consisting of geometric forms varying in 6 attributes with 2 levels of each, (b) sequence displays, consisting of 6 plus and/or minus signs in a row. The arrangement of cards in the stimulus displays was ordered or random. The principal results were: (a) 2-person groups used the focusing strategy more, required fewer card choices to solution, and required more time than individuals; (b) form displays resulted in more use of the focusing strategy than sequence displays, with no difference in number of card choices; (c) no difference between ordered and random arrays in use of the focusing strategy or number of card choices.

Bruner, Goodnow, and Austin (1956) distinguish the two basic selection strategies of focusing and scanning in concept attainment. In focusing *S* tests the relevance of all the possible hypotheses involved in a particular attribute or attributes by choosing a card differing in one (conservative focusing) or more (focus gambling) attributes from a positive focus card. In scanning he tests specific hypotheses, either singly (successive scanning) or all at once (simultaneous scanning) or some intermediate number. In general, focusing is a more successful strategy, which Bruner et al. (1956) interpret as due to the more difficult memory requirements of scanning. This interpretation is supported by later experiments demonstrating the important effects of memory in concept learning by Cahill and Hovland (1960) and Hunt (1961). In a study of individual vs. group memory, Perlmutter (1953) found better memory for groups. From these considerations of memory effects in concept learning and in individuals vs. groups, the present study hypothesized that two-person cooperative groups would (*a*) use the focusing strategy more than individuals, and (*b*) solve concept-attainment problems in fewer card choices than individuals. If memory is a relevant variable, individuals should have more repetitions of card choices, more hypotheses, more repetitions of hypotheses, and more untenable hypotheses than groups.

Concept-attainment research typically has used stimulus displays with simple geometric forms such as triangles or squares, varying in several attributes such as color or size. These displays present *S* with one unified perceptual object having several attributes of variability. In applying a focusing strategy, attributes are successively abstracted from the unified object and the set of hypotheses involved tested for applicability to the concept. The *S* may thus be set to use a focusing strategy because of the nature of the stimulus display. However, if the attributes were not in a single perceptually unified figure, but rather in a series of separate figures, *S* might perceive the situation as a problem in the permutations of discrete objects and thus be set to use a scanning strategy by directly testing specific hypotheses. Thus, the present experiment used two types of stimulus displays: (*a*) the typical geometric forms with six attributes, (*b*) a sequence of six plus signs and/or minus signs in a row. It was hypothesized that the form displays would result in more use of the focusing strategy than sequence displays, and thus in fewer card choices.

---

[1] This paper is based on a dissertation submitted to the Department of Psychology of Northwestern University in partial fulfillment of the requirements of the PhD degree. The author is greatly indebted to D. T. Campbell, C. P. Duncan, W. F. Hill, and B. J. Underwood for their very helpful advice and criticism.

METHOD

Design.—A 2 × 2 × 2 factorial design was used with the following variables: (*a*) number of persons (individual or two-person group), (*b*) stimulus display (form or sequence), (*c*) array (ordered or random).

Subjects.—Twelve male introductory psychology students were randomly assigned to each of the four individual conditions and 24 to each of the four group conditions.

Stimulus displays.—The problem materials were 24, 28 × 44 in. white posterboards, each containing an 8 × 8 array of 64, 2½ × 4 in. cards drawn in colored ink with dark outlines. The 64 cards represented all possible combinations of six attributes with two levels of each.

Form displays consisted of the following attributes and values: (*a*) shape: square or triangle (*b*) size: large or small (*c*) number: one or two (*d*) color: red or green (*e*) pattern: striped or solid (*f*) borders: one or two. Sequence displays consisted of all combinations of six plus and/or minus signs in a row. In order to facilitate reference to the six positions, each was a different color, so that the color name was the attribute and plus or minus the value of each color. The attributes and values were listed on a reference card which *S* could use throughout the experiment. All cards were numbered from 1 to 64 by rows in the upper right corner.

The stimulus displays were arranged in six ordered arrays, in which the position of each card varied systematically in relation to the other cards, and six random arrays. For example, on the first ordered array the top four rows of cards had one border and the bottom four rows two. Each attribute occupied a different position on each of the six arrays. Within each treatment condition two replications were randomly assigned to each of the six arrays, and within each array one replication was randomly assigned to two-attribute and one to three-attribute concepts.

Problems.—All *Ss* within form or sequence displays received the same five problems in the same order. The two-attribute form display problems were: (*a*) striped, triangle (*b*) small, solid (*c*) one figure, two borders (*d*) two figures, large (*e*) green, square. Three-attribute problems added the values (*a*) large (*b*) square (*c*) large (*d*) red (*e*) one border, to the corresponding two-attribute problem. The two-attribute sequence display problems were: (*a*) green minus, red plus (*b*) black minus, yellow minus (*c*) orange plus, red plus (*d*) blue minus, green plus (*e*) blue plus, black plus. Three-attribute problems added the values (*a*) yellow plus (*b*) blue minus (*c*) green minus (*d*) orange minus (*e*) red plus, to the corresponding two-attribute problem.

Procedure.—The instructions explained the meaning of conjunctive concepts and the nature of the task, pointed out the attributes and values and the ordered or random arrays, and emphasized that the problems were to be solved in as few card choices as possible, regardless of time (Laughlin, 1964).

RESULTS

The means for the eight treatment groups for focusing strategy, number of card choices to solution, and time to solution over all five problems are given in Table 1. Focusing strategy was scored according to three rules: (Rule 1) Each card choice had to obtain information on one new attribute. New information was obtained if the card choice altered only one attribute not previously proven irrelevant (conservative focusing) or, if more than one attribute was altered (focus gambling), the instance was either positive or the ambiguous information correctly resolved on the

TABLE 1

Mean Number of Card Choices to Solution, Focusing Strategy, and Time to Solution over Five Problems

	Form Displays		Sequence Displays	
	Ordered	Random	Ordered	Random
Individual				
Focusing Strategy	3.48	3.20	2.89	2.50
Card Choices to Solution	36.83	32.58	34.08	34.67
Time in Minutes	16.50	19.20	12.60	15.20
Group				
Focusing Strategy	4.19	3.88	3.62	3.38
Card Choices to Solution	26.08	30.42	26.58	25.08
Time in Minutes	22.70	27.00	20.10	29.10

Note.—Maximum focusing strategy is 5.00.

next card choice by altering only one attribute. (Rule 2) If a hypothesis was made it had to be tenable considering the information available. Untenable hypotheses were of two types: (*a*) a hypothesis for a value of an attribute when the other value had previously occurred on a positive instance, e.g., the hypothesis "red square" when a green instance had been positive; (*b*) a hypothesis for a value which had previously occurred on a negative instance, e.g., the hypothesis "red square" when an instance with a red square had been negative. (Rule 3) Neither the card choice nor hypothesis could be a repetition of a previous card choice or hypothesis. Each card choice and accompanying hypothesis (if given) that satisfied these three rules was counted as an instance of focusing, and the total number of such instances was divided by the total number of card choices to give a continuous focusing score from .00 to 1.00.

Two-person groups used the focusing strategy more than individuals at the .001 level, F (1, 88) = 23.02, and required fewer card choices to solution at the .001 level, F (1, 88) = 14.37. Groups required more time to solution at the .001 level, F (1, 88) = 17.13. Form displays resulted in more use of the focusing strategy than sequence displays at the .001 level, F (1, 88) = 14.07. There was no difference between form and sequence displays on number of card choices to solution, F (1, 88) < 1. The differences between ordered and random arrays on number of card choices to solution and focusing strategy were not significant, F (1, 88) < 1 and F (1, 88) = 3.88, respectively. None of the interactions of any variable were significant.

The means for the eight treatment groups for number of repetitions of card choices, hypotheses, repetitions of hypotheses, and untenable hypotheses over all five problems are given in Table 2.

Individuals had more repetitions of card choices than groups at the .05 level, F (1, 88) = 5.78; more hypotheses at the .001 level, F (1, 88) = 13.09; more repetitions of hypotheses at the .05 level, F (1, 88) = 4.40; and more untenable hypotheses at the .01 level, F (1, 88) = 7.99. None of the interactions of any variable were significant.

Since the six ordered and six random arrays and the two-attribute vs. three-attribute concepts were essentially control variables of secondary

TABLE 2

Mean Number of Repetitions of Card Choices, Hypotheses, Repetitions of Hypotheses, and Untenable Hypotheses over Five Problems

	Form Displays		Sequence Displays	
	Ordered	Random	Ordered	Random
Individual				
Repetitions of Card Choices	2.08	4.83	1.77	5.57
Hypotheses	15.08	19.92	14.00	16.67
Repetitions of Hypotheses	1.33	2.42	.83	1.83
Untenable Hypotheses	3.83	5.58	2.83	2.75
Group				
Repetitions of Card Choices	1.08	2.08	1.08	.67
Hypotheses	11.75	12.75	11.67	10.00
Repetitions of Hypotheses	.92	1.33	.42	.33
Untenable Hypotheses	1.58	2.25	1.58	.33

interest, they were analyzed across all conditions by a one-way analysis of variance rather than as orthogonal effects in the factorial design. There were no significant differences between either the six ordered arrays or the six random arrays for either focusing strategy or number of choices, $F\ (5, 42) < 1$ for all four. There were thus no gross board-position effects. There was no significant difference between two-attribute and three-attribute concepts on number of choices, $F\ (1, 94) < 1$, but three-attribute concepts resulted in more use of focusing at the .025 level, $F\ (1, 94) = 6.38$.

As hypothesized, two-person groups required fewer card choices to solution and used the focusing strategy more than individuals. However, as in all individual vs. group problem solving, the objection may be made that this result is merely an artifact, due to the greater probability of one person in each pair in group conditions being above the overall mean in ability. If these persons dominate the problem solving, group superiority is merely an artifactual function of their greater ability. In order to test this objection, the correction model of Taylor and McNemar (1955) was used. Two-person "groups" were formed by randomly pairing all the 48 *S*s in individual conditions, and giving each pair the score obtained by the better of the two individuals. The procedure was repeated with another random pairing to obtain another 24 scores, and the 48 scores were compared with the actual group scores by a one-way analysis of variance. This analysis showed no difference between actual and artificial groups for number of card choices to solution, $F\ (1, 46) < 1$. The actual groups used the focusing strategy more than artificial groups at the .001 level, $F\ (1, 46) = 23.37$.

Discussion

The difference between two-person groups and individuals on number of card choices to solution was not upheld with the Taylor and McNemar (1955) correction model, so the most parsimonious conclusion is that it is merely an artifactual function of the better individual in each pair. However, the greater use of the focusing strategy by groups was upheld with the correction model, so that two persons working together used the strategy more than they would have working alone. This seems to be due to better memory for the groups, as indicated by significantly fewer repetitions of card choices, hypotheses, repetitions of hypotheses, and untenable hypotheses for groups than individuals. In addition, the opportunity for discussion in the groups may have aided them in the realization that focusing is a strategy designed to reduce memory requirements and insure a constant increment of information on each card choice.

As hypothesized, form displays resulted in more use of focusing than sequence displays. This seems to reflect a difference in the perception of the two displays. The form displays would seem to be perceived as unified figures with several attributes of variation, which may set *S* to use focusing by successively abstracting attributes and testing the set of hypotheses involved in each attribute. The sequence displays would seem to be perceived as a series of different discrete objects, which may set *S* to consider the problem in terms of the permutations of many objects. Thus, he would tend to use a scanning strategy in order to test specific hypotheses directly. However, despite the difference in the use of focusing, the hypothesis that form displays would require fewer card choices to solution was not supported.

The greater use of focusing with three-attribute concepts than with two-attribute concepts is only suggestive because it was based on a one-way analysis of variance rather than a factorial effect. However, it does reflect the relative efficiency of focusing vs. scanning for the two types of concepts. With focusing, three-attribute concepts can be solved in a maximum of four choices, while two-attribute concepts require a maximum of five. On the other hand, scanning is more difficult with three-attribute concepts, because, given the initial positive focus card, there are 20 possible three-attribute concepts and only 15 possible two-attribute concepts.

Contrary to the results of Bruner et al. (1956), there were no differences in the use of focusing or number of card choices between ordered and random arrays. Since random arrays only complicate an already complex situation, it would seem better to limit future research on strategies to ordered arrays.

REFERENCES

BRUNER, J. S., GOODNOW, J. J., & AUSTIN, G. A. *A study of thinking.* New York: Wiley, 1956.

CAHILL, H. E., & HOVLAND, C. I. The role of memory in the acquisition of concepts. *J. exp. Psychol.*, 1960, **59**, 137–144.

HUNT, E. B. Memory effects in concept learning. *J. exp. Psychol.*, 1961, **62**, 598–604.

LAUGHLIN, P. R. Speed versus minimum-choice instructions in concept attainment. *J. exp. Psychol.*, 1964, **67**, 596.

PERLMUTTER, H. V. Group memory of meaningful material. *J. Psychol.*, 1953, **35**, 361–370.

TAYLOR, D. W., & MCNEMAR, O. W. Problem solving and thinking. *Annu. Rev. Psychol.*, 1955, **6**, 455–482.

(Received June 2, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 3, 328–331

GENERALIZATION TO SIMILAR AND OPPOSITE WORDS

RICHARD K. OLSON[1]

University of Oregon

30 Ss were taught a lever-pressing response to 7 stimulus words which elicit "opposites" and 7 stimulus words which elicit "similars" as primary responses in free-association tests. Generalization was tested to the primary responses. It was expected that generalization would be greater for "similars" since it seemed that they should be a greater part of the meaning response to the stimulus. Significant generalization occurred to both opposites and similars ($p < .001$), but there was no significant difference between opposites and similars ($p > .05$). The hypothesis was rejected.

Classical-conditioning approaches have used implicit meaning responses as a basis for semantic generalization. Razran (1949) has shown that a salivation response may be conditioned to words and generalization obtained to their synonyms. Since the conditioned response was based on similarity in meaning and not the appearance of the word, it seems that the conditioned response is contingent upon the internal meaning response to the stimulus word. Mink (1963) used a voluntary lever-pressing response to study generalization as related to forward and reverse association. His results indicated that a generalization effect can be obtained by pairing a voluntary response with words and testing to associates of the words.

Karwoski and Schachter (1948) found that attempts to classify free-association responses led to a set of categories that are essentially degrees of similarity, and of contrast or difference. Responses involving identification are included in a basic category of "similars" as opposed to those that are "opposites." It is possible that response words of different categories are related to the meaning response of the stimulus words in strengths which are not measured by frequency of association.

A method based on Mink's (1963) generalization study has been devised to measure the relation of the response word to the stimulus word at the implicit level, and without the usual association request. With this method, associations in Karwoski's two basic categories, similarity and opposition, were compared. It was expected that more generalization would occur between similars than between opposites of equal free-association frequency.

METHOD

Subjects.—Sixteen men and 14 women were used. They were volunteers from introductory psychology classes at the University of Oregon.

Materials.—All words used in the experiment were selected from the Minnesota word-association response norms (Russell & Jenkins, 1954), with the exception of two that were taken from Deese's (1962) unpublished norms for 500 words. Stimulus words and their primary responses were classified into categories of "opposites" and "similars" by two judges. The judges agreed on 85% of the pairs. Those pairs on which there was disagreement were rejected.

Each S-R pair in the opposites category was matched with a similar S-R pair that had approximately the same frequency of forward association. Reverse-association data for each pair was obtained from Deese's norms. Seven matched pairs were selected so that the

---

[1] The writer wishes to express appreciation to Barry Anderson for his assistance in the preparation of this research, and to Fred Attneave for his suggestions and comments.

TABLE 1

FREQUENCIES OF ASSOCIATION AND GENERALIZATION

Stimulus Word	Response Word	Frequency of Forward Association (Base 1000)	Frequency of Reverse Association (Base 50)	Frequency of Generalization to Response Word (Base 180)
SHORT	TALL	397	30	43
EATING	FOOD	390	30	100
COLD	HOT	348	16	62
ANGER	MAD	353	20	35
BEAUTIFUL	UGLY	209	4	44
COMMAND	ORDER	196	11	75
ROUGH	SMOOTH	439	22	53
CARPET	RUG	460	6	45
EARLY	LATE	500	17	55
HAMMER	NAIL	537	23	32
HATE	LOVE	340	8	99
OCEAN	WATER	314	2	73
BITTER	SWEET	652	4	75
BLOSSOM	FLOWER	672	2	86

mean frequencies of reverse association for each category were approximately equal. The pairs balanced for forward and reverse association are given in Table 1 with their respective frequencies. No response word was related to a stimulus word other than its own by an association frequency greater than 1% of the total. Also 17 neutral words were chosen which had no associative relationship greater than 1% to any of the words in the 14 S-R pairs.

The learning list was made up from the 14 stimulus words with three of the neutral words (to be designated as "dummy" words) at each end of the list to help control for serial-position effects. The opposite and similar stimulus words were alternated in the series.

The test list consisted of the 14 response words, 5 dummy words used in the learning list, and 11 neutral words. The words were distributed randomly, with the restriction that no more than 2 words from a single classification could appear in succession. A second form of the learning and test lists was made up in which the positions of opposites and similars were exchanged to counterbalance for possible position effects. Any obvious physical similarity among the words was avoided.

Apparatus.—The lists were presented on a Hull-type memory drum enclosed in a wooden box. Two one-way windows, 1 × 2 in., each exposing one side of the drum, were set in the front of the box. A small light bulb was set on the inside of the box next to each window. A list could be viewed through a window only when its respective light was on. The words were typed in capitals on white kymograph paper. Protruding from the front of the box was a spring lever which *S* held between his thumb and forefinger. The lever could be pressed down and was automatically returned to its original position. The response lever was connected to a counter so that the number of responses could be recorded.

Procedure.—The *S* was seated in front of the apparatus and was told that this was an experiment in recognition time and reaction speed. The *S* was presented with two lists of words. The first list appeared in the left window of the apparatus. The *S* was told to press the lever quickly when each word appeared and to remember the word as the lever was pressed.

The learning list was presented twice with a 16-sec. break between presentations. The interval between all words was about $\frac{1}{16}$ of a second. The speed of presentation was one word per 1 sec. on the learning list and one word per $1\frac{1}{2}$ sec. on the test list.

After the second presentation of the learning list there was a 2-min. break. During this time *S* was told that the test list would contain most of the words in the learning list. Each time *S* thought he recognized a word from the learning list, he was to press the lever as quickly as possible. Response to even the faintest recognition was emphasized. The test list was 30 words long and was presented continuously six times.

The *E* was seated behind *S* and recorded lever-pressing responses for each word. The first form of the learning and test lists was presented to the first 15 *S*s and the second

TABLE 2

Analysis of Variance of Categories

Source	*df*	*MS*	*F*
Categories (C)	2	10.01	27.3
S	29	1.554	—
S × C	58	.367	—

form to *S*s 16–30. After each *S* had completed the test, he was asked if he recognized any peculiar relationship between the test list and the learning list.

Results

Total generalization scores for each response word are given in Table 1. A mean score for each *S* in each classification was obtained by adding the number of responses and dividing by the number of words in the category. There was an average over *S*s of 2.043 responses per word for "opposites," 2.103 for "similars," 4.253 for "dummy," and 1.075 for neutrals.

Analysis of the results was made in terms of analysis of variance and the t test for correlated measures. The results of analysis of variance are summarized in Table 2.

Analysis of variance for the three categories yielded $F\,(2, 58) = 27.3$. Two t tests were performed to determine the significance of the difference between similars and opposites and between opposites and neutrals. Generalization to similars was not significantly greater than to opposites. A comparison yielded $t\,(29) = 0.36$. Comparison of opposites and neutrals yielded $t\,(29) = 8.06$ ($p < .001$).

Additional tests were performed to determine the correlation between frequency of forward association and frequency of generalization and between reverse association and generalization. An $r = .05$ ($p > .05$) was obtained for forward association and $r = -.24$ ($p > .05$) for reverse association.

In reply to the question concerning the relationship between the learning list and the test list, three *S*s said they saw a relationship of similars and opposites. The other 27 *S*s did not report noticing the relationship.

Discussion

The way in which generalization is mediated in this design may be clarified by Fig. 1. The stimulus word, response word, and motor response are represented by the letters A, B, and C, respectively. Letters in parentheses indicate implicit meaning responses. In the first stage, the connection between A and B is assumed to be established in *S*'s background. In the second stage the stimulus word, A, is paired with a motor response, C. In the third or test stage, generalization is tested to the response word, B.

With the meaning response acting as a mediator, generalization may occur in the following manner: In Stage 2, A elicits a meaning response of which the response word, B, is a part. The implicit B is then connected to C. In the test stage, B may elicit an implicit A as part of its meaning response, and this may facilitate the response, C. Thus either the B-C connection established in Stage 2, or the implicit A elicited in the test stage, or both, serve to mediate generalization.

Only a very small and insignificant difference occurred between similars and opposites. The hypothesis that greater generalization would occur to similar responses must be rejected. At first glance these results are rather surprising.

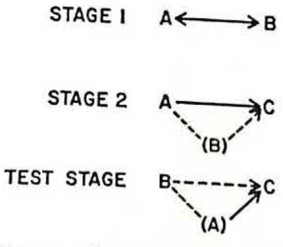


Fig. 1. Stages of mediated generalization.

It seems that a response which is similar in meaning to the stimulus word would be a greater part of the meaning response than a response which is opposite in meaning. If this reasoning is carried to its extreme, there should be no generalization at all to opposites. If the opposite response is completely different in meaning from the stimulus, it is difficult to see how mediation could take place through the meaning response. Deese (1962) has suggested an explanation that fits these results. He pointed out that opposites are adjectives, ". . . and they are descriptive of a common state modified only in a kind of secondary way." For example, the words hot and cold both describe temperature. The contrast is in the degree of temperature. It is possible that the common dimension described by two opposing adjectives is the mediating link in the meaning response.

The semantic relations described above by no means exhaust the possible bases for mediation. There are speech-habit relations and many others which are not semantic. Different *S*s may use different modes and different methods may bring out different relationships among words. Carroll, Kjeldergaard, and Carton (1962) have found that individuals may have different tendencies for giving opposites. Ervin (1963) has suggested that the explicit association of opposites is based on erroneous implicit substitutions resulting from ambiguous anticipations.

The low and insignificant correlations between generalization and free-association frequency are probably due to the relatively limited range of associative frequency. Also speech habits and number of competitors for response may be factors confusing this relationship.

In conclusion, there appears to be very little, if any, difference in generalization for similars and opposites. It is likely that semantic relations among words are augmented by other relations to provide the mediating link in generalization.

REFERENCES

Carroll, J. B., Kjeldergaard, P. M., & Carton, A. S. Number of opposites versus number of primaries as a response measure in free association tests. *J. verbal Learn. verbal Behav.*, 1962, **1**, 22–30.

Deese, J. E. Unpublished norms for responses to 500 words, unpublished manuscript, Johns Hopkins University, 1962.

Deese, J. E. On the structure of associative meaning. *Psychol. Rev.*, 1962, **69**, 161–175.

Ervin, S. M. Correlates of associative frequency. *J. verbal Learn. verbal Behav.*, 1963, **1**, 422–431.

Karwoski, T. F., & Schachter, J. Psychological studies in semantics: III. Reaction times for similarity and difference. *J. soc. Psychol.*, 1948, **28**, 103–120.

Mink, W. D. Semantic generalization as related to word association. *Psychol. Rep.*, 1963, **12**, 59–67.

Razran, G. H. S. Semantic and phonetographic generalization of salivary conditioning to verbal stimuli. *J. exp. Psychol.*, 1949, **39**, 642–653.

Russell, W. A., & Jenkins, J. J. The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Word-Association Test: Studies in the role of language in behavior. Technical Report No. 11, 1954, University of Minnesota, Contract N8-onr-66126.

(Received June 8, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 3, 332–333

SUPPLEMENTARY REPORTS

ACOUSTIC FACILITATION OF VISUAL DETECTION[1]

WILLIAM H. WATKINS[2] AND CARL E. FEEHRER

Decision Sciences Laboratory, Electronic Systems Division, Bedford, Massachusetts

11 Os were required to judge which of 4 temporal intervals contained a visual signal, in an experiment involving a total of 10,900 trials. Under some conditions, potentially useful time information was conveyed by accompanying sound stimulation, while it was lacking under others. Highest detectability of the signal was associated with an acoustic condition having white noise bursts coincident with each observation interval. Those detection scores were significantly superior to a "reciprocal" condition having the identical amount of acoustic time-specification information. Detection was poorest under continuous noise and silence, which were not discernably different in their effects. Simple time cueing was inferred not to provide an adequate explanation for the results.

An earlier forced-choice study (Watkins, 1964) found that detectability of a light signal increased when a recurrent pattern of noise bursts was substituted for continuous noise. The pattern responsible for this effect consisted of bursts coincident with the times of possible visual signal occurrence. The facility used did not permit inclusion of two sound conditions of potential interest: silence, and noise continuous except for silent periods coincident with observation intervals. In the absence of data regarding performance in silence, no appraisal of the absolute degree of facilitation accompanying introduction of noise was possible. More importantly, it was not possible to determine whether the increased detectability noted under the recurrent noise pattern was due simply to the provision of more precise cues to times of possible signal occurrence or to intrinsic properties of the pattern.

In the present experiment, interest centered upon the adequacy of observation-time-cueing as a basis for explaining differences in performance.

Method.—The *O* was required to judge which of four temporal observation intervals (OIs) contained a visual signal. He registered his decision by depressing one of four "vote" buttons, connected to a tape punch.

The apparatus used in this experiment is described elsewhere (Watkins, Nickerson, & Schjelderup, 1964). The elements were an observer station, control equipment, and a set of computer programs. Auditory and visual displays and a special *O*'s chair were contained within an anechoic chamber. The visual display panel was located 11 ft. in front of *O* and consisted of a transilluminated panel with attached time-cueing, "ready," and "vote" lamps. The brightness of the panel was 24 footlamberts (ftl.). A PDP-1 digital computer was used to generate signal sequences and to analyze data.

Figure 1 shows (I) the four OIs and (II) one trial sequence. The significance of "C" and "S" segments associated with four component periods of each trial was as follows: C: The segment during which a $2\frac{1}{2} \times 3\frac{1}{4}$ in. cue panel, at lower left corner of the main display area, bearing the letter L, was lighted. S: The segment during which the signal could have been presented—i.e., an OI.

A signal consisted of an increase of the order of 4% in the brightness of a 1.3-in. fixation spot (circular, 25-ftl. patch) in the center of the display panel. It was produced by raising the current flow through a projector lamp by a given amount (determined by signal strength needs) for 165 msec.

Four independent blocks of 12 trials comprised a run of 48 successive trials. Signals were distributed among the four OIs such that each block included three occurrences of the signal in each OI. For any *O*, a run sequence was used only once.

Four men served as the first sample of *O*s; seven women as the second. All were naive concerning the purposes of the investigation.

After brief practice, each *O* in the first

---

[1] This is Electronic Systems Division Technical Report 65-156 of the Air Force Electronic Systems Division, Air Force Systems Command. This research was performed at the Decision Sciences Laboratory as part of Project 7682, Man-Computer Information Processing. Further reproduction is authorized to satisfy the needs of the United States Government.

[2] Major, United States Air Force.

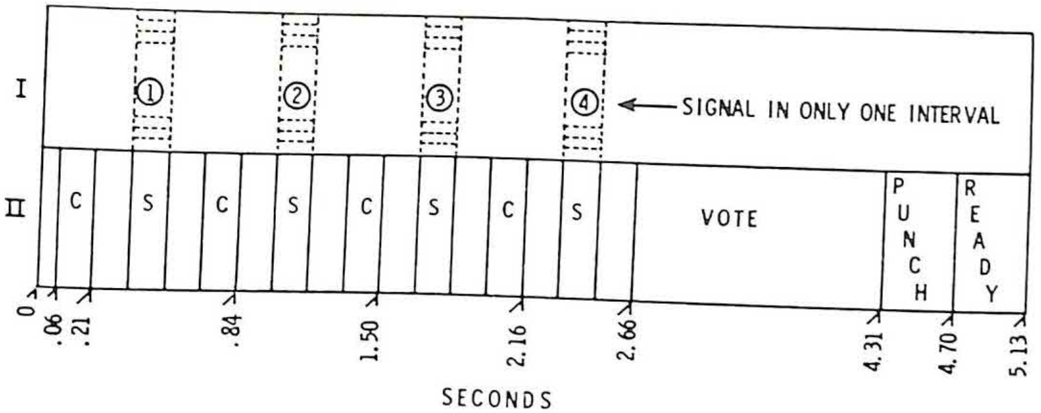


FIG. 1. Standard time configuration of trials. (I: Four possible signal intervals. II. Basic time periods. Between the .06- and 2.66-sec. points, there were 16 equal segments of approximately 165 msec.)

sample performed eight sets of four runs with each set consisting of one run under each of four acoustic conditions. The duration of each run was 4 min.; a rest period followed. Acoustic conditions used were (Cond. 1) continuous noise; (Cond. 2) noise continuous except for interruptions coincident with observation intervals; (Cond. 3) silence except for noise bursts coincident with observation intervals; and (Cond. 4) silence.

Since silence had proved essentially equivalent to continuous noise for the first sample (see results), *E*s elected to terminate "silent" testing at this point, in the interest of collecting more data under conditions of greater concern. Accordingly, the second sample was tested under Cond. 1, 2, and 3 only. After practice, these *O*s accomplished a total of 33 sets (i.e., 99 runs). The noise (presented through earphones) had a frequency range of 100–6800 cps and a sound pressure level (SPL) of 70 db. overall. Knowledge of results was provided in the form of numbers of trials correct at the end of each run.

Results.—For the first sample, group means (percentage of correct trials) were: Bursts at OIs: 46.6; Interruptions at OIs: 43.8; Continuous noise: 40.6; Silence: 40.0. For each of the *O*s, Bursts at OIs was the condition showing highest detection rate. That condition was significantly superior to the other intermittent noise condition: $F(1, 3) = 13.5$, $p < .05$. The data of the second sample yielded means of 62.3 (Bursts), 56.4 (Interruptions), and 49.3 (Continuous noise). The t tests comparing these (correlated) means revealed that all conditions differed from the others significantly ($p < .01$).

Discussion.—The evidence provided by the first sample showed that continuous presentation of the moderate level of noise chosen did not establish a state of elevated or depressed visual sensitivity.

Time-cueing alone appeared inadequate as a mechanism to account for the superior detectability of visual signals found when OIs were accompanied by noise bursts. If bursts had served simply to mark the boundaries of OIs, thus informing *O* as to when he should watch for a signal, then the reciprocal condition—interruptions at OIs—should have served as well. The latter condition was inferior to the bursts condition but better than continuous noise. Since OI-phased noise presentation, even of the "interruptions" variety, did aid visual detection, the provision of increased time certainty may not be rejected as a partial explanation. Additional relevant studies (Watkins & Feehrer, 1964) corroborate these conclusions. However, the precise role of noise bursts in this type of task remains unknown.

REFERENCES

WATKINS, W. H. Effect of certain noises upon detection of visual signals. *J. exp. Psychol.*, 1964, **67**, 72–75.

WATKINS, W. H., & FEEHRER, C. E. Investigations of acoustic effects upon visual signal detection. *USAF ESD TR*, 1964, No. 64-557.

WATKINS, W. H., NICKERSON, R. S., & SCHJELDERUP, J. R. An automated facility for forced-choice signal detection experimentation. *USAF ESD TDR*, 1964, No. 64-383.

(Early publication received January 25, 1965)

Journal of Experimental Psychology
1965, Vol. 70, No. 3, 334–335

ISOLATION EFFECTS WHEN PAIRED ASSOCIATES ARE PRESENTED SERIALLY[1]

SLATER E. NEWMAN AND G. ALFRED FORSYTH
University of North Carolina, Raleigh — *Purdue University*

This experiment studied the effect of isolating one or both terms of a pair on learning the pair. The pairs were presented in the same order on each trial, the pair with the isolated term(s) appearing near the middle of the list. Isolation led to faster learning of the pair, but did not affect the rate of learning the entire list.

Erickson (1963) has reported that when one or both terms of an already isolated pair are further isolated learning of the pair is facilitated. Isolating both terms of the pair produced greater facilitation than isolating only the response term.

This experiment studied the effects of isolating one or both terms of a pair (which had not already been isolated) on learning the pair. The pairs in the list were presented in the same order on all trials. As in serial-learning experiments (e.g., Smith & Stearns, 1949) the isolated term (s) appeared near the middle of the list, eighth in a list of 14 pairs.

Method.—The *S*s were 120 males with normal color vision enrolled as undergraduates at the University of North Carolina at Raleigh. None had served previously in a learning experiment. They were randomly assigned to the 12 treatments, through use of a balanced Latin square.

Twenty-eight disyllables from Noble's (1952) list, each with accent on the first syllable, were used to make up the 14-pair list. The mean "*m*" value for the stimulus terms was 6.70 and for the response terms 6.67. The range of *m* values was 3.21–9.43 for stimulus terms and 3.36–9.61 for response terms. No two stimulus terms began with the same letter, nor did any two response terms. There were four conditions of isolation—None (N), Stimulus Term (S), Response Term (R), and Both Terms (S-R). Isolation was produced by typing one or both terms in the eighth (critical) pair in red on pairing trials and, when the stimulus term was to be isolated, (i.e., in Groups S-R and S) on test trials as well. All other terms were typed in black. Three different orders of the pairs (X, Y, and Z) were used under each isolation treatment. The critical pairs and their *m* values are HUNGER (6.02)-CAPTAIN (6.88), QUARTER (5.98)-INSECT (7.39), and ZEBRA (7.12)-EFFORT (5.13).

A pairing-test procedure was used, each of the 15 pairing trials being followed by a test. A 2-sec. rate was used for both pairing and test trials. The interval between trials was 4 sec. The *S*s were instructed not to attend to the order of the pairs during training. They were also told that at the end of training, they would be tested in an order different from any used during training. After the fifteenth test, *S*s were given two more tests. Half were tested in the same order as during training and then in a different order (Cond. S); the other half were tested in the different order first (Cond. D). Each *S* was then tested for color vision, given a postexperimental interview, and dismissed.

Results and discussion.—The number correct during training for the critical pair was determined for each *S* and a 4×3 analysis of variance (Isolation × List Order) applied to these data. Only the effect of Isolation was significant, $F\ (3, 108) = 12.06$, $p < .001$. Duncan's test indicated that the mean for each of the three isolation groups was higher ($p < .01$) than the mean for Group N. No other differences were significant though the difference between Groups S-R and R approached significance ($.10 > p > .05$). As in Erickson's (1963) experiment, the S-R mean exceeded the R mean. The means are presented in Table 1. Most *S*s gave the response for the critical item correctly (i.e., to its stimulus term) the first time they gave it; the percentage of *S*s in each group was S-R, 90; S, 97; R, 97; and N, 77.

To determine whether isolation affected the rate of learning the list, the trial on which the entire list was given correctly was recorded for each *S* and a Kruskal-Wallis test done on these data. *H* was not significant ($p > .05$). An analysis of variance for the total number of correct responses during training also showed the *F* for Isolation not to be significant ($p > .05$).

---

[1] Based on an MS thesis by the junior author at the University of North Carolina at Raleigh. This report was prepared by the senior author. We appreciate the comments of R. L. Erickson. Support for this study was provided by Office of Naval Research Contract Nonr 486(08). Reproduction of this article in whole or in part is permitted for any purpose of the United States Government.

TABLE 1

MEANS FOR TOTAL CORRECT FOR CRITICAL PAIR AND ENTIRE LIST AND MEDIAN NUMBER OF TRIALS TO ONE PERFECT TRIAL

Group	N[a]	Total Correct (Critical Pair) M	Total Correct (Entire List) M	First Perfect Trial Mdn.
S-R	30	13.5	162.5	8.0
S	30	13.5	174.0	7.0
R	30	12.2	163.2	8.0
N	30	10.3	165.1	7.5

[a] Data for Orders X, Y, and Z have been combined for each treatment.

Two types of errors were studied: (*a*) the number of times that the response term from the critical pair was given instead of other response terms from the list and (*b*) the number of times that other response terms from the list were given in place of the response term for the critical pair. For each *S* both scores were divided by the total number of errors involving response terms from the list, and Kruskal-Wallis tests performed on each of these variables. Neither *H* was significant ($p > .05$).

For *S*s who had the same order on Test 16 as during training (Cond. S) the mean was 13.68 on Test 15 and 13.52 on Test 16; for *S*s tested in a different order (Cond. D) the means were 13.67 and 13.08. For each *S* the difference between Test 15 and Test 16 was determined. A *t* test was used to compare the means of these differences between Cond. S and D, t (118) = 2.52. Though *t* was significant, the small absolute difference between the two treatments combined with the high mean on Test 16 for Treatment D suggests that at the end of training *S*s were responding to the stimulus term and not to the cue for serial position for most of the pairs.

The results of this experiment indicate that isolating one or both terms of a pair facilitates learning the pair. These results accord with those obtained by Erickson (1963). Neither rate of learning the list nor type of error were, however, affected by isolation.

REFERENCES

ERICKSON, R. L. Relational isolation as a means of producing the von Restorff effect in paired-associate learning. *J. exp. Psychol.*, 1963, **66**, 111–119.

NOBLE, C. E. An analysis of meaning. *Psychol. Rev.*, 1952, **59**, 421–430.

SMITH, M. H., JR., & STEARNS, E. The influence of isolation on the learning of surrounding materials. *Amer. J. Psychol.*, 1949, **62**, 369–381.

(Received May 15, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 3, 335–337

ON THE DIFFERENT EFFECTS OF RANDOM REINFORCEMENT AND PRESOLUTION REVERSAL ON HUMAN CONCEPT IDENTIFICATION[1]

SOLON B. HOLSTEIN AND DAVID PREMACK

University of Missouri

Levine's random reinforcement (RR) procedure was replicated using 6 as well as 2 dimensions. Contrary to a suggestion by Bower and Trabasso, RR produced as marked a decrement with 6 dimensions as with 2. However, a backward-learning curve showed post-RR learning to be essentially stationary. Both the RR decrement and the stationarity can be handled by the Bower-Trabasso model on the assumption that RR increases H, the number of S's hypotheses.

In standard two-category concept-identification experiments, Bower and Trabasso (1963) have shown that *S*s whose S-R assignments were reversed after the occurrence of every second error performed as well as controls that were never reversed. These data support the cue-selection theories of Restle (1962) and Bower and Trabasso (1963), for the effect of an error in these theories is to return *S* to the state he was in before he began work on the problem, so that a change in the S-R assignment after an error should have no effect.

However, Levine (1962) has shown that random reinforcement (RR), unlike presolu-

---

[1] This research was carried out while the first author was on a National Aeronautical and Space Administration predoctoral research fellowship. The research was aided by Grant NSF GB 1379 from the National Science Foundation and by Grant MH-05798 from the National Institutes of Health.

tion reversal, does have a decremental effect, though it should not according to the cue-selection theories, for both RR and presolution reversal involve changing the S-R assignment after *S* makes an error. Nevertheless, Levine has shown that starting *S* out with no more than six RRs before shifting him (without signal) to a consistent S-R assignment has a clear decremental effect.

Trabasso and Bower (1964) have attributed Levine's outcome to his use of only one or two irrelevant dimensions in contrast to the six which they have used. According to their theory, *S* attends selectively to cues from the stimulus cards and tests hypotheses regarding the relevance of the cues. If *S*'s response is correct, it is supposed that he continues to use the same hypothesis; otherwise, he resamples at random from the set of hypotheses, and further, resamples with replacement so that his set of hypotheses remains constant in size. Bower and Tra-

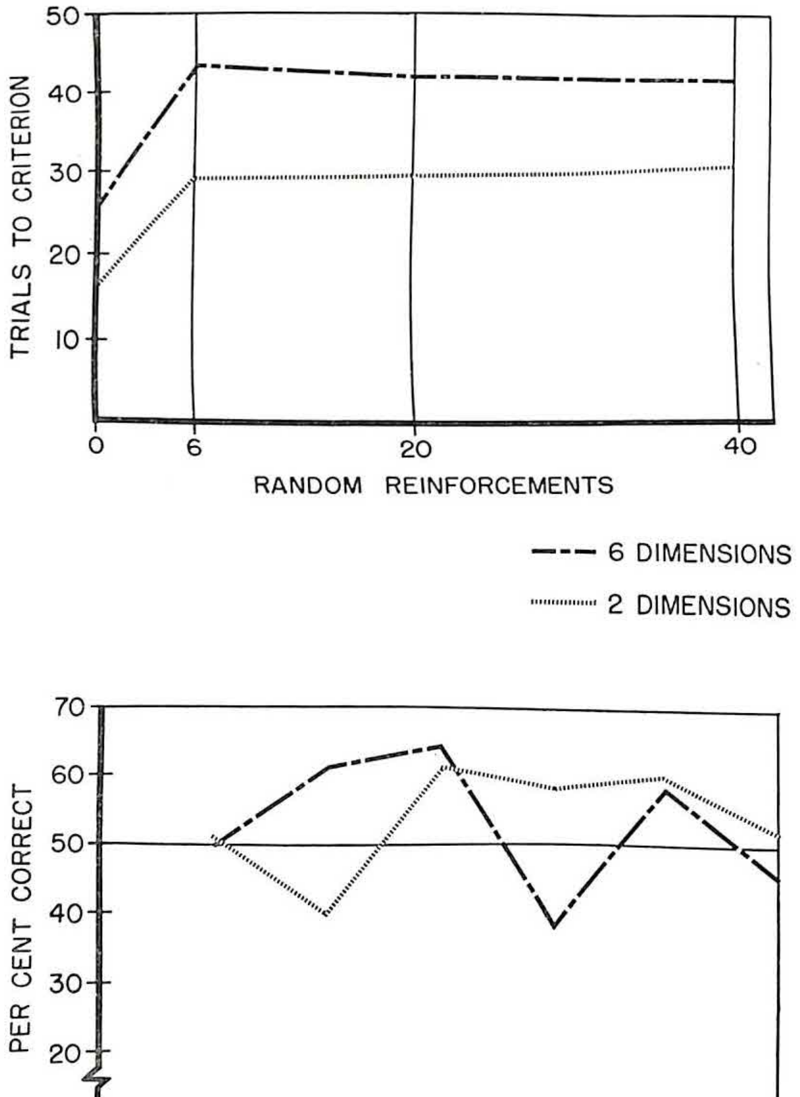


FIG. 1. Trials to criterion as a function of number of random reinforcements with separate curves for the two- and six-dimension groups (upper panel); and backward learning curve showing percentage of success prior to the last error plotted over successive blocks of five trials (lower panel).

basso suggest that when Levine used only one or two irrelevant dimensions, his *S*s did not resample with replacement; rather, they were able to remember the hypotheses that were disconfirmed during the RR phase. The present experiment is a test of the Bower-Trabasso interpretation; Levine's RR procedure is replicated with six as well as two irrelevant dimensions.

Method.—The *S*s were 160 University of Missouri students who were assigned to 16 groups and tested in individual sessions. The *S*'s task was to learn to classify cards into two categories, VEC and NONVEC, on the basis of a single relevant dimension. Number of dimensions (2 vs. 6), number of RRs (0, 6, 20, 40), and two types of instructions (vague vs. explicit) were combined factorially to form a 2 × 4 × 2 design. In the vague instructions *S*s were told only that the cards could be sorted into VEC and NONVEC, and that *E* would tell him when his choice was "right" or "wrong"; *S*s given the explicit instructions were, in addition, shown two cards that contained all values of the six dimensions, the two values of each dimension were pointed out to *S*, and *S* was told that the problem had a simple solution.

Three decks of cards were used. The dimensions were form (diamond vs. square), color (black vs. white), size (large vs. small), direction of a diagonal passing through the figure (right vs. left), parallels bordering the figure (horizontal vs. vertical), and background (striped vs. nonstriped). All 64 possible combinations of the six dimensions were used to construct a deck of 90 cards in which the different combinations were distributed randomly throughout the deck with the restriction that no one value of a given dimension occurred more than twice in succession. Two different two-dimension decks were used to check the effect of different irrelevant dimensions; color and form, and color and size.

Training consisted of two phases. In the first and RR phase, *S*s were told "right" or "wrong" after each choice according to a predetermined random schedule for either 0, 6, 20, or 40 trials, after which, without signal, all *S*s were consistently reinforced on the basis of color, i.e., black-VEC, white-NONVEC. The cards were shown for either 50 trials or until *S* made 15 correct responses in a row.

Results.—The upper panel of Fig. 1 shows that, contrary to the suggestion by Bower and Trabasso, the six-dimension case was no less subject to impairment by RR than the two-dimension case. The deficit in both cases appears to be maximal by at least 6 RR, which is in agreement with Levine's (1962) data. An analysis of variance of the trials to criterion showed that the main effects of RR and number of dimensions were both significant, $F\,(3.144) = 21.62$, $p < .01$; $F\,(1.44) = 45.85$, $p < .01$, respectively; and that neither the main effect of instructions nor any of the interactions approached the 5% level. Results for the color-form and color-size decks did not differ and were, therefore, combined in the analysis.

The lower panel shows a backward-learning curve for the two- and six-dimension groups. The essential stationarity prior to the last error was shown further in the vincentized data; percentage of success for the first and second halves of the trials prior to the last error was 58 and 55, respectively, for the two-dimension case, and 49 and 49, respectively, for the six-dimension case.

The stationarity data support the Bower-Trabasso model, while the RR impairment in the six-dimension case appears to contradict the model. However, an interpretation of the effect of RR can be shown that is wholly in keeping with the Bower-Trabasso model. So far the only changes in *S*'s hypotheses (H) that have been considered are (*a*) that *S* samples with replacement so that H remains constant, or (*b*) that *S* remembers disconfirmed hypotheses, does not sample with replacement, so that H decreases. But consider a third alternative: Error-producing operations such as RR lead *S* to form new hypotheses, thus increasing the size of H. In terms of the cue-selection theories, an increase in H by RR would have exactly the desired effect, viz., a lowering of the probability of learning after RR with no effect upon the all-or-none character of learning. That error-producing operations may, in fact, not only increase H but also change the character of *S*'s hypotheses is being examined in studies aimed at getting at the composition of H.

REFERENCES

BOWER, G., & TRABASSO, T. Reversals prior to solution in concept identification. *J. exp. Psychol.*, 1963, **66**, 409–418.

LEVINE, M. Cue neutralization: The effects of random reinforcements upon discrimination learning. *J. exp. Psychol.*, 1962, **63**, 438–443.

RESTLE, F. The selection of strategies in cue learning. *Psychol. Rev.*, 1962, **69**, 329–343.

TRABASSO, T., & BOWER, G. Presolution reversal and dimensional shifts in concept identification. *J. exp. Psychol.*, 1964, **67**, 398–399.

(Received May 18, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 3, 338–341

EFFECTS OF LIST LENGTH IN THE EBBINGHAUS DERIVED-LIST PARADIGM[1]

ROBERT K. YOUNG, DAVID T. HAKES, AND R. YALE HICKS
University of Texas

Previous experiments on the derived-list (DL) serial-learning paradigm have left undetermined whether the paradigm yields positive or negative transfer. They have also used unusually long (16-item) lists. The present experiment investigated the 1st-order DL paradigm using 8-, 12-, and 16-item lists. Significant negative transfer was obtained for the 12- and 16-item conditions throughout learning. For the 8-item condition, early negative transfer gave way to slight positive transfer late in learning. Reasons for the differences as a function of list length were considered. It was concluded that the 1st-order DL paradigm yields negative transfer.

Hakes, James, and Young (1964) recently presented evidence that the derived-list (DL) serial-learning paradigm does not yield positive transfer. Studying first-, second-, and third-order DL conditions, they found positive transfer early in second-list learning only. This tended to dissipate rapidly and be replaced by negative transfer which was, however, significant only for the first-order condition.

The lists used were 16 items in length. The nature of the paradigm dictates the use of such extraordinarily long lists if DL conditions higher than the first-order are to be studied. The question arises, however, of the generality of the findings obtained using long lists. The present experiment was designed to investigate this question and to provide further evidence on the amount and direction of transfer in the DL paradigm.

Method.—Serial lists of 8, 12, and 16 items were employed. For each list length, Experimental *S*s learned two lists related in a first-order derivation while Control *S*s learned two unrelated lists. Two independent item orders of each length were constructed from each of two independent item sets. Two *E*s each tested half the *S*s in each condition, yielding a 3 × 2 × 2 × 2 × 2 factorial design. Ninety-six *S*s served in the experiment ($N = 2$ per cell) and were randomly assigned to conditions.

Lists for each list-length cnodition were constructed from two samples of 16 adjectives according to the procedures followed by Hakes et al. The testing procedures of the earlier experiment were also followed.

Results and discussion.—For the first list, trials to a criterion of one errorless trial were analyzed. List Length was the only significant effect, $F\ (2,\ 48) = 64.17$, $p < .001$, indicating that speed of learning was inversely related to list length. The lack of a significant effect for Treatments, $F\ (1,\ 48) = 1.85$, supports the hypothesis of equivalent groups.

For the second list, mean numbers of trials to reach successive criteria are presented in Fig. 1. Trials to a criterion of one errorless trial were analyzed for this list. Significant effects were obtained for List Length, $F\ (2,\ 48) = 26.73$, $p < .001$, for Treatments, $F\ (1,\ 48) = 8.14$, $p < .01$, and for the List Length × Treatments interaction, $F\ (2,\ 48) = 3.20$, $p < .05$. No other effect reached significance.

The Treatments effect indicates that overall negative transfer was obtained. For the 12- and 16-item conditions, respectively, 49% and 44% negative transfer was obtained. For the 8-item condition, 13% positive transfer was obtained. An orthogonal comparison of the 8-item condition with the other two conditions was significant, $F\ (1,\ 48) = 6.31$, $p < .05$, indicating that the amount of transfer for the 8-item condition differed from that for the other two conditions. This difference is related to the significant List Length × Treatments interaction.

Transfer early in second-list learning was analyzed for numbers of trials to reach a criterion of two correct responses. List Length was not a significant effect, $F\ (2,\ 48) = 0.56$, but Treatments differed, $F\ (1,\ 48) = 7.53$, $p < .01$. Figure 1 indicates that negative transfer was obtained early in learning for all List-Length conditions. This result differs from that obtained by Hakes

---

[1] The present research was supported by National Science Foundation Grant GB-979. The authors wish to express their appreciation to Carlton T. James and Jeanne W. Brown for their assistance in the collection of the data.

et al., who found early positive transfer. In both cases, however, the Treatment differences are small, attaining significance because of extremely small within-group variances.

Serial-position curves for the Treatments within each List-Length condition are presented in Fig. 2. These indicate that the negative transfer for the 12- and 16-item conditions occurred primarily in the middle of the lists, with some tendency for positive transfer

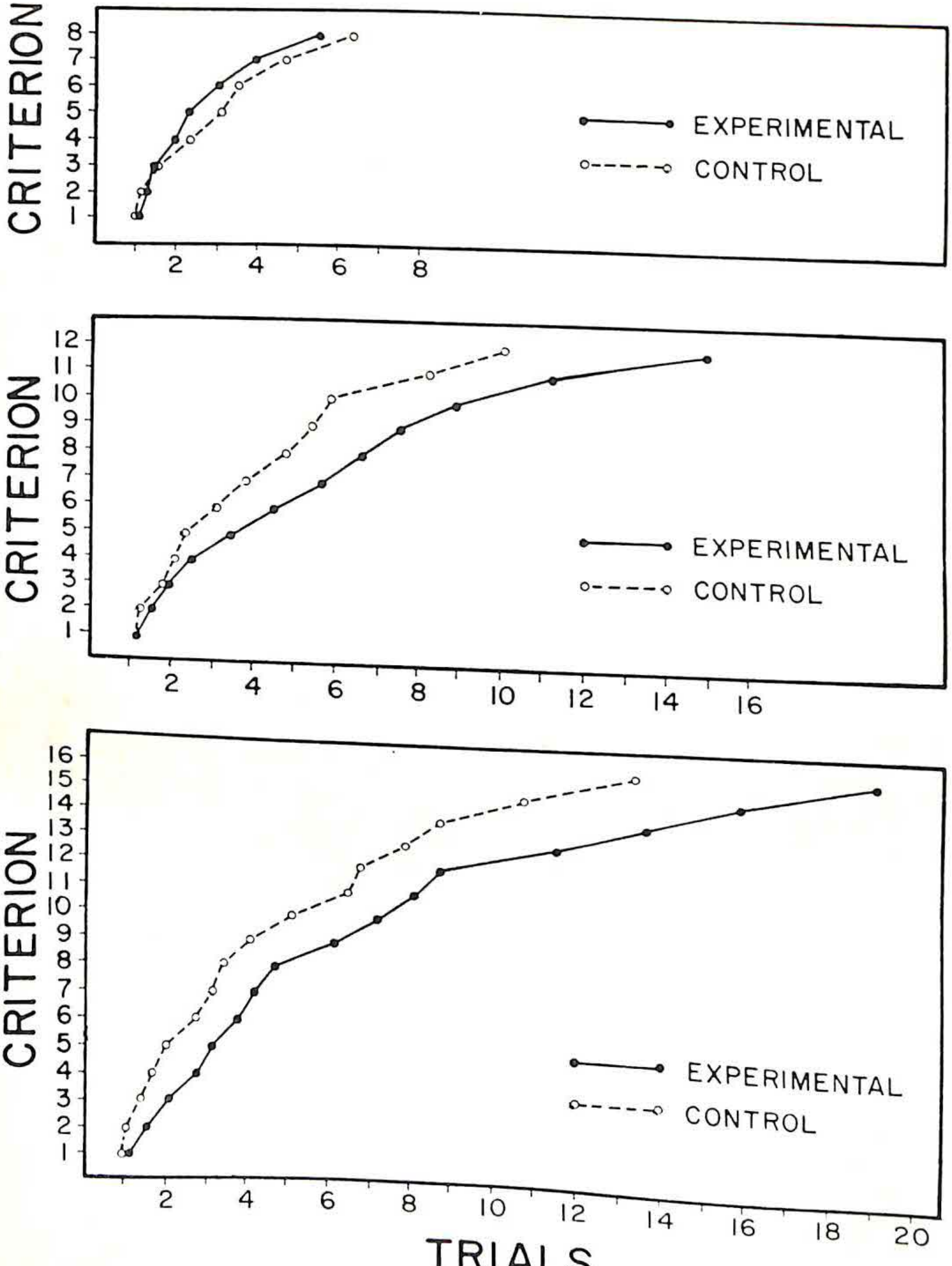


FIG. 1. Mean numbers of trials to successive criteria during second-list learning.

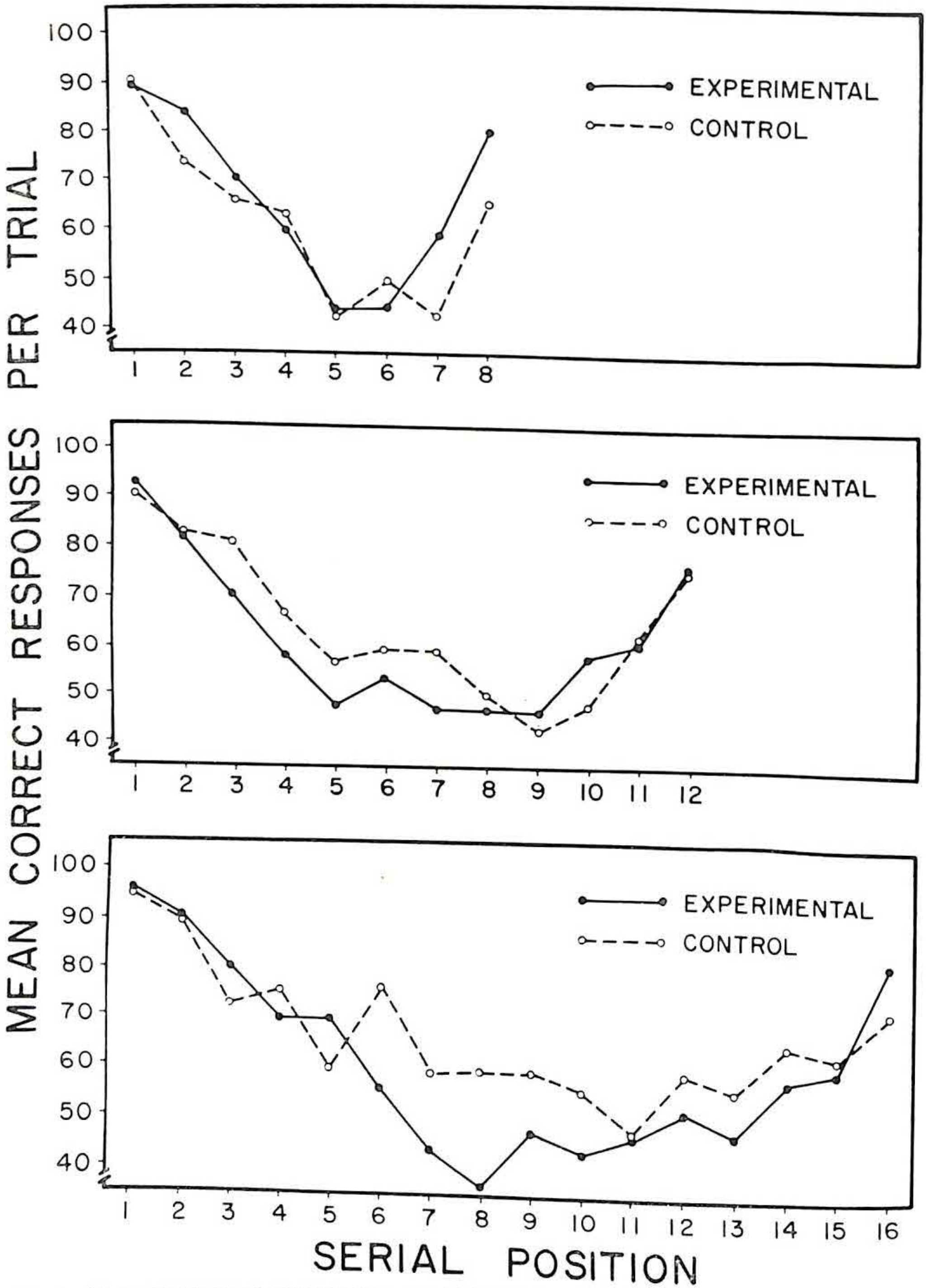


FIG. 2. Mean percentages of correct responses per trial at each serial position during second-list learning.

at the ends. The 8-item condition also yielded positive transfer at the ends of the list, but items in the middle failed to yield marked negative transfer. No statistical technique is available which permits comparison of these trends across lists of unequal length. It appears, however, that for the 12- and 16-item conditions, the Experimental-Control comparisons show considerably greater depression of Experimental group performance than does the same comparison for the 8-item condition.

Consistent with the Hakes et al. data, it may be concluded that the first-order DL paradigm yields negative transfer. Two interpretations of the 8-item condition's failure to yield negative transfer suggest themselves: (*a*) it is possible that with short lists the relationship between lists is apparent to some *Ss*; or (*b*) the items at the ends of the list, which tend to yield positive transfer for all list lengths, comprise a greater proportion of the list for the 8-item condition than for either of the other two.

The present results are consistent with both ordinal position and sequential-association hypotheses of serial learning. The former predicts that the stimulus (position) which, in the first list, is associated with one response is, in the second list, associated with a different response, constituting an A-B, A-B_{jumbled} negative transfer paradigm. The latter predicts that for the first list to be learned, the direct associations must be stronger than any remote associations. During second-list learning, these formerly correct associations are incorrect and should be a source of interference. Thus the present data do not permit a differential test of current serial-learning hypotheses. It is apparent, however, that under a variety of conditions the first-order DL paradigm yields negative transfer.

REFERENCE

HAKES, D. T., JAMES, C. T., & YOUNG, R. K. A reexamination of the Ebbinghaus derived-list paradigm. *J. exp. Psychol.*, 1964, 68, 508–514.

(Received May 18, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 3, 342

Beginning with the first issue of Volume 69 and continuing through the remainder of that volume and Volume 70 (1965), the titles and authors of accepted papers will be listed here following Supplementary Reports. It is being supported on an experimental basis by the APA Project on Scientific Information Exchange in Psychology, and at the end of the year, the outcome of this trial will be evaluated and consideration given the advisability of continuing the listing.

This listing plus those published in the preceding 1965 issues are a portion of the backlog of manuscripts accepted by this journal. Such listing will allow readers to become aware of research many months in advance of journal publication.

Manuscripts Accepted for Publication in the *Journal of Experimental Psychology*

Compensatory Hue Shift in Simultaneous Color Contrast as a Function of Separation between Inducing and Test Fields: Tadasu Oyama and Yun Hsia*: Psychology Department, 370 Schermerhorn Hall, Columbia University, New York, New York 10027.

Stimulus Similarity and the Effect of Reinforcement in a Pseudo-Concept-Identification Task: Juliet P. Shaffer* and Robert K. Remple: Department of Psychology, University of Kansas, Lawrence, Kansas.

A Figural Aftereffect Produced by a Phenomenal Dichotomy in a Uniform Contour: Bernard Weitzman*: New School for Social Research, 66 West 12th Street, New York, New York.

Verbal Behavior and Problem Solving: Some Effects of Labeling in a Functional Fixedness Problem: Sam Glucksberg* and Robert W. Weisberg: Department of Psychology, Green Hall, Princeton University, Princeton, New Jersey 08540.

Associative Asymmetry as a Function of Pronounceability: Clifton W. Gray* and Slater E. Newman: Norwich Hospital, P. O. Box 508, Norwich, Connecticut 06316.

Initial Transfer in the Mediation of Paired Associates: Charles Clifton, Jr.*: Institute of Child Behavior and Development, University of Iowa, Iowa City, Iowa.

Directional Asymmetry of Motion Aftereffect: Thomas R. Scott,* Abraham D. Lavender, Ronald A. McWhirt, and Donnie A. Powell: Veterans Administration Hospital, Columbia, South Carolina.

Transition Probability, Word Order, and Noun Abstractness in the Learning of Adjective-Noun Paired Associates: Igor Kusyszyn* and Allan Paivio: Department of Psychology, University of Western Ontario, Middlesex College, London, Ontario, Canada.

Differentiation versus Unlearning of Verbal Associations: Norman J. Slamecka*: Department of Psychology, Waterman Building, University of Vermont, Burlington, Vermont 05401.

Effects of Formal Similarity on Cue Selection in Verbal Paired-Associate Learning: Jean C. Cohen* and Barbara S. Musgrave: Department of Psychology, Smith College, Northampton, Massachusetts.

Whole versus Part Learning of Paired-Associate Lists: Leo Postman* and Judith Goggin: Institute of Human Learning, University of California, 2241 College Avenue, Berkeley, California 94720.

Recovery of Conditioned UCR Diminution Following Extinction: M. C. Morrow*: Division of Social Sciences, University of South Florida, Tampa, Florida 33620.

Effects of CS Omission Following Avoidance Learning: J. M. Bloom and Byron A. Campbell*: Department of Psychology, Princeton University, Princeton, New Jersey 08540.

Unlearning in Two Directions: Isabel M. Birnbaum*: Institute of Human Learning, University of California, 2241 College Avenue, Berkeley, California 94720.

---

* Asterisk indicates author for whom address is supplied.

JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR

invites you to enter your subscription and to receive, free of charge as part of your subscription, reference cards for all articles appearing in the volume. These standard 3 × 5 cards provide subject headings, author, title, bibliographic citation, and abstract. Space is also available for the reader to enter his own classification. Because of the enthusiastic reception by subscribers, cards are being prepared for all back volumes of JEAB.

It is not too late to enter a subscription for Volume 8 (1965).

Major articles appearing in the July 1965 issue:

Shettleworth, Sara and Nevin, John A. **Relative rate of response and relative magnitude of reinforcement in multiple schedules.**

Rohles, Frederick H., Jr. **Wind as an aversive stimulus.**

Dardano, J. F. **Modification of observing behavior.**

McMillan, D. E. and Patton, R. A. **Differentiation of a precise timing response.**

Cole, James Lawrence. **Force gradients in stimulus generalization.**

Ferster, C. B. and Hammer, Clifford. **Variables determining the effects of delay in reinforcement.**

Ferraro, D. P., Schoenfeld, W. N., and Snapper, A. G. **Sequential response effects in the white rat during conditioning and extinction on a DRL schedule.**

Zeiler, Michael D. **Solution of the intermediate size problem by pigeons.**

Technical Notes

Tentatively scheduled for September 1965:

Beasley, Julaine L. and Seldeen, Barbara L. **The effect of acceleration on food-reinforced DRL and FR.**

Fisher, G. L. **Saline preference in rats determined by contingent licking.**

Hake, D. F. and Azrin, N. H. **Conditioned punishment.**

Kaplan, Michael, Jackson, Bruce, and Sparer, Richard. **Escape behavior under continuous reinforcement as a function of aversive light intensity.**

Kintsch, Walter. **Frequency distribution of interresponse times during VI and VR reinforcement.**

Thomas, John R. **Discriminated time-out avoidance in pigeons.**

Wertheim, George A. **Behavioral contrast during multiple avoidance schedules.**

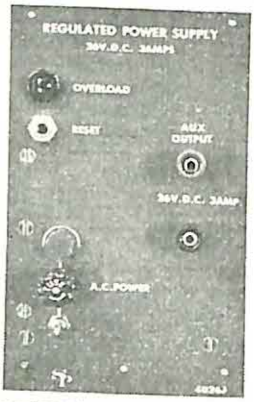
Technical Notes

Subscriptions: $8.00 students; $9.50 individuals (personal use only); $16.00 institutions. Subscriptions are entered on a volume (calendar year) basis only.

Please continue to address all orders to the Business Office: Kay Dinsmoor, Department of Psychology, Indiana University, Bloomington, Indiana 47405.

SCIENTIFIC PROTOTYPE

POWER
SUPPLIES



REGULATED POWER SUPPLIES FOR REED RELAY AND RESEARCH EQUIPMENT

Here are two power supplies that are designed specifically for use with all relay programming and control equipment.

The units are designed for use with Scientific Prototype's new line of Reed Relay Equipment.

Extremely high reliability is insured by the all-silicon semi-conductor circuitry. No routine maintenance, including fuses, is required. SP's power supplies feature good line and load regulation, excellent ripple and transient load performance, and complete short circuit protection—all at a price directly competitive to unregulated or semi-regulated units.

SP's 4026-J and 4027-J Regulated Power Supplies combine excellent performance with moderate cost and are fully compatible with existing 24 to 28 VDC relay equipment. The 4026-J Power Supply provides 26 VDC @ 3A in a ¼ rack clip-on. The 4027-J is rated at 26 VDC @ 10A in a 5¼" x 19" rack panel. For complete catalog information, write today.

Behavior Analysis Apparatus . . . operant conditioning kits, rodent test cages, pellet dispensers, soundproof chambers, recorders, noise generators, shock sources, timers, switches, digital test equipment and special systems.

SCIENTIFIC PROTOTYPE MFG. CORP.
615 W. 131 Street, New York, N. Y. 10027
Telephone: (212) 368-6000

SCIENTIFIC PROTOTYPE

NEW PLANT · NEW OFFICES

Please note our new address and phone number. Our new, larger plant enables us to offer an even wider range of products and services.

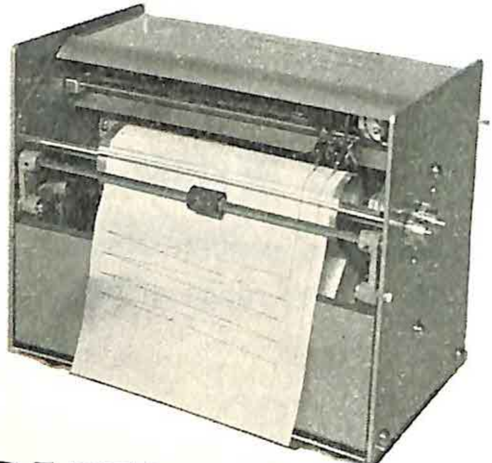
NEW PRODUCTS · NEW CATALOGS

Scientific Prototype provides an unusually broad line of standard equipment:

Rodent test cages, pellet dispensers, soundproof chambers, recorders, noise generators, shock sources, timers, audio switches, two and three channel tachistoscopes, cumulative recorders, power supplies, a complete line of reed relay logic equipment, a complete line of solid state programming equipment, closed circuit T.V., plus equipment custom built to your specifications.

A COMPLETE 75-PAGE REED RELAY EQUIPMENT CATALOG IS NOW AVAILABLE (LETTERHEAD REQUEST ONLY).

SCIENTIFIC PROTOTYPE MFG. CORP.
615 W. 131 Street, New York, N. Y. 10027
Telephone: (212) 368-6000



OUR MODEL

C-3

AND

C-3, S.H.S.

CUMULATIVE RECORDERS

FIRST AND FOREMOST FOR OVER 12 YEARS OF CONTINUED USE IN LABORATORIES THROUGHOUT THE WORLD. (AT NO INCREASE IN PRICE)

TRULY:

- **ACCURATE**
- **RELIABLE**
- **DURABLE**

Ralph Gerbrands Company

Scientific Instruments

8 Beck Road
Arlington, Mass. 02174

Tel. 648-6415
Area Code 617

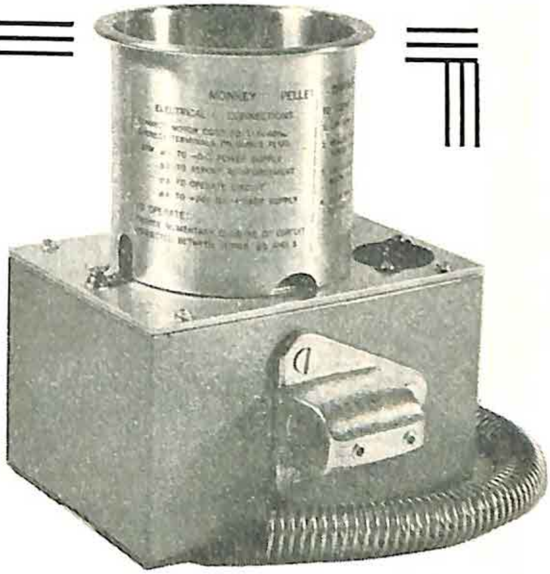
A NEW

A.C. MOTOR DRIVEN

Dispenser

FOR

Monkey Pellets
Marbles
M & M Candies



FEATURES

- **INSTANT DELIVERY** OF PELLET—**NO LATENCY**
- **LOW** D.C. CURRENT DRAIN
- HIGH **RELIABILITY**
- LOW NOISE LEVEL
- **DURABLE**-LONG LIFE CONSTRUCTION
- CAN BE MODIFIED TO DELIVER ONE OR TWO PELLETS ON COMMAND
- WRITE OR PHONE FOR DETAILS NOW

Ralph Gerbrands Company

Scientific Instruments

8 Beck Road
Arlington, Mass. 02174

Tel. 648-6415
Area Code 617

$10 a year Bimonthly $2.00 a copy

Journal of
Educational
Psychology

Studies of
learning
teaching
measurement

Order from:
AMERICAN PSYCHOLOGICAL
ASSOCIATION
1200 Seventeenth Street, N. W.
Washington, D. C. 20036

Journal of
Experimental Psychology

VOL. 70, NO. 4 OCTOBER 1965

REINTERPRETATION OF ONE FORM OF BACKWARD AND FORWARD MASKING IN VISUAL PERCEPTION [1]

CHARLES W. ERIKSEN AND JAMES F. COLLINS

University of Illinois

2 experiments were performed on the masking effect of a ring stimulus upon forced-choice recognition of letter forms when the ring was presented concurrently with the form or preceding or following the form at delay intervals up to 250 msec. 10 Ss were employed in Experiment I (backward masking) and 10 in Experiment II (forward masking). Masking effects on recognition accuracy were obtained for delay intervals as long as 80 msec. Maximum masking occurred when ring and form occurred concurrently. The masking functions obtained for forward and backward masking appeared highly similar if not identical. An interpretation was suggested in terms of visual perceptual persistence where masking is due to the greater difficulty of perceiving a form surrounded by a black ring. Possible perceptual mechanisms for the persistence were considered.

If a briefly presented visual stimulus is followed by a second stimulation within a brief time interval, seldom longer than 100 msec., there are circumstances under which the first signal will not be perceived by an *S* even though the duration of its exposure was long enough for perception to have occurred had the second stimulus not been presented. Alpern (1952) provided a description of some of these phenomena and Raab (1963) has contributed a more recent review.

The backward masking, or as it is sometimes called, "metacontrast," has tended to be treated as a single phenomenon, but recent evidence has made increasingly clear that several phenomena are involved with different explanations or underlying processes (Eriksen & Collins, 1964; Eriksen & Hoffman, 1963). Eriksen and Hoffman (1963) have advanced an explanation for certain cases of masking in terms of a temporal luminance summation between the two stimulus fields with a resulting reduction in contrast ratio for the masked figure. The luminance-summation, contrast-reduction hypothesis has received strong support from several studies (Eriksen & Hoffman, 1963; Eriksen & Steffy, 1964; Steffy & Eriksen, 1965) and most recently Thompson (in press) has shown that an empirically

---

[1] This investigation was supported by Public Health Service Research Grant MH-1206 and a Public Health Service Research Career Program Award K6-MH-22,014.

determined function relating perceptual recognition of forms to contrast ratios, predicts quite precisely the recognition accuracy for forms masked by a succeeding second-field luminance where the effective contrast ratio for the form was computed, assuming luminance summation.

While the luminance-summation hypothesis seems to explain quite well certain instances of backward masking, there are other circumstances where masking is obtained that cannot be attributed to luminance summation and resulting contrast reduction. Kolers (1962) and Kolers and Rosner (1960) have reported masking effects on a disk stimulus followed by a second stimulus consisting of a ring under circumstances where the adapting and delay fields were illuminated at a level that would have reduced luminance summation and contrast reduction to minimal levels. Also Averbach and Coriell (1961), using quite bright adapting and delay fields, found impaired recognition for letters when the letter was circumscribed by a ring occurring later in time. Unfortunately the methodology employed in these studies was not very rigorous. The Kolers (1962) study was defective in the use of the duration of disk exposure as a dependent variable and along with the Kolers and Rosner (1960) experiment provided no control for shift in *S*'s criterion of report (Swets, Tanner, & Birdsall, 1961). Indeed, Kolers and Rosner (1960) found that the probability of masking at a single delay interval *and identical stimulus conditions* varied from less than 20 to greater than 70% depending upon the range of delay intervals *S* was judging in that particular session. While the authors interpret this in terms of anchoring effects, it most likely reflects a changing criterion on the part of the *S* correlated with the range of stimuli being judged. If *S* is making judgments over a long range of delay intervals, he experiences very little masking on presentations involving longer delays and is apt to adopt a very rigorous criterion for reporting the presence of the disk. When judging over a shorter range of delays, particularly over those intervals where masking is maximal, *S* is apt to adopt a more lenient criterion of disk presence since on none of the exposures does he get a very clear perception of the disk. The result is that any masking phenomenon is itself masked by a variable and unmeasured false-alarm rate.

In the present experiments we have attempted to remedy past defects in methodology by: (*a*) obtaining a more sensitive measure of masking through the use of a recognition criterion for forms; (*b*) the use of a forced-choice psychophysical procedure with confidence ratings on judgments; (*c*) a constant-stimulus exposure time adjusted to each *S*'s perceptual capability; and (*d*) observations of the effect on perceptual recognition when form stimulus and masking ring are presented concurrently.

EXPERIMENT I

The purpose of this experiment was to map the retroactive interference produced by a ring stimulus upon recognition of simple forms when the ring stimulus occurred concurrently with the form or at delay intervals as much as 250 msec. after termination of the form stimulus. Two groups of *S*s were run. The Group A *S*s were run first at delay intervals designed to map the general outline of the interference function. On the basis of data obtained from

Group A, Group B *S*s were run at intervals designed to provide maximum information in the critical regions of the masking or interference function.

Method

Subjects.—Ten adult college students volunteered to serve and were paid a fixed sum for the experiment. The five *S*s in Group A contained four males and the five *S*s in Group B one male. All *S*s were trained over several practice sessions that involved perceptual recognition but not masking phenomena. All had normal or corrected to normal visual acuity.

Procedure.—A three-field tachistoscope, previously described (Eriksen & Hoffman, 1963), was used to present the stimulus sequences. An adaptation field with a luminance of .572 apparent footcandles (ftc.) contained a black H subtending .2° of angle which was used as a fixation point. Each of the two stimulus fields had a luminance of .200 apparent ftc. In order to minimize confounding from luminance-summation, contrast-reduction effects, the adaptation field remained on continuously and exposures of the two stimulus fields, F 1 and F 2, were superimposed upon the adaptation-field luminance. Thus, at all delay intervals between the two stimulus fields the adaptation-field luminance was present.

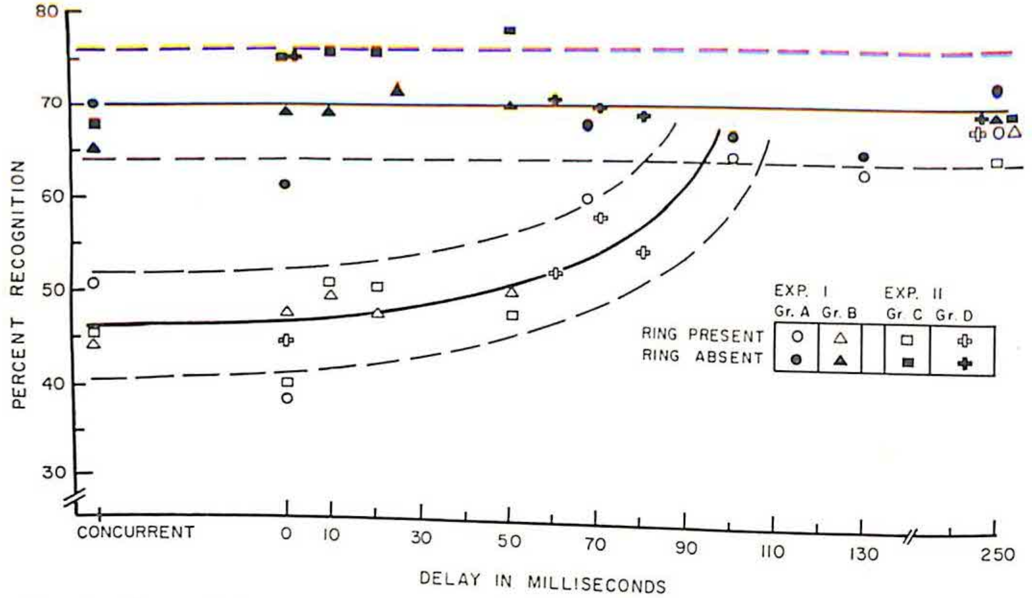
The stimulus forms consisted of the capital letters A, T, and U, presented as black figures against a white ground and each subtending a maximum of .2° of angle. Previous research had found these letters to be approximately equally recognizable and equally confusable, one with another, in tachistoscopic presentations. The ring stimulus was also black on a white ground and had an outer diameter of .67° and an inner diameter of .30°. The ring was always presented in F 2 and so located as to center the form occurring in F 1.

Prior to beginning the experimental sessions each *S* engaged in two practice sessions of approximately 1 hr. each. During these practice sessions *S*'s forced-choice recognition function for the three stimuli was determined when the letters were presented without the occurrence of the ring-masking stimulus. Instead only a blank white card was flashed at 0-delay interval in F 2 after presentation of a letter in F1. From these practice judgments a base-level duration of exposure was determined for each *S* that yielded approximately 70% forced-choice correct recognition. This base level was then used for this *S* as the exposure duration for the forms in F 1. The mean value of this base-level duration was 44.0 msec. with a range from 27 to 57 for the 10 *S*s. The ring stimulus when it occurred in F 2 always had a constant exposure time of 100 msec. for all *S*s.

Each *S* served for five experimental sessions of approximately 40-min. duration during which he made 108 judgments as to which letter had been presented. The *S* was instructed before each stimulus sequence to fixate clearly the fixation point and when it appeared sharply and clearly to add pressure to the trigger which initiated the program.[2] After each trial *S* reported whether A, T, or U had occurred along with a rating of his subjective confidence on a 3-point scale. The *S*s in Group A each made 90 judgments at each of five delay intervals (0, 68, 100, 130, and 250 msec.) and 90 for the concurrent situation. The *S*s in Group B differed only in that the delay intervals were: concurrent, 0, 10, 25, 50, and 250 msec. There were two kinds of trials for each delay interval as well as for the concurrent presentation. On half the trials, randomly selected, the letter form was followed by the ring stimulus after the appropriate delay (ring-present condition). On the other half, only a blank white card in the second field was flashed after the appropriate delay (ring-absent condition). By presenting half of the trials without the ring stimulus, a constant control was maintained on recognition accuracy without the presence of interference and also a control for a possible distracting effect of a flicker in the second stimulus field was provided. On the concurrent presentations the letter stimulus and ring were on the same stimulus card and were presented in F 1 followed at 0 delay by the 100-msec. flash of the blank white card in F 2. The *S*s made judgments at all delay intervals, and the concurrent situation during each experimental session and the order of delay intervals within sessions were counterbalanced across sessions and across *S*s.

---

[2] We have found in past work (Eriksen & Hoffman, 1963) that this method of permitting *S* to initiate the sequence by manipulating the switch or trigger much as a rifleman squeezes off a shot, results in very stable recognition functions for tachistoscopically presented stimuli. It succeeds in eliminating variations in accommodation which account for considerable variability in tachisotoscopic presentations.



	EXP. I Gr. A	EXP. I Gr. B	EXP. II Gr. C	EXP. II Gr. D
RING PRESENT	○	△	□	✚
RING ABSENT	●	▲	■	✚

FIG. 1. Recognition as a function of delay interval between the letter form and the ring stimulus or blank field. (The dotted-line curves indicate the + and − 2σ range around the solid-line curves.)

Results

The number of correct letter recognitions independent of subjective confidence level was analyzed by a three-way analysis of variance (ring present vs. absent, delay interval, and *S*s).[3] Since different delay intervals were employed, a separate analysis was required for each of the two groups.

For both groups the effect of the ring-present–ring-absent conditions was significant beyond the .001 level and the effect of delay interval was significant beyond the .025 level. The interaction between ring-present–ring-absent and delay interval was also significant for both groups ($p < .025$). None of the interactions involving the *S* variable were significant but there was a significant main effect attributable to difference in *S*s ($p < .025$). This was anticipated due to inability to equate precisely all *S*s at the 70% recognition point on base-level duration.

In Fig. 1, percentage of correct recognition of the forms has been plotted for each of the two groups as a function of ring-present vs. absent conditions and delay interval. The data points for the ring-absent condition have been fitted by a straight line of 0 slope. For the ring-present or masking condition the data points of 100-msec. delay and less have been fitted to the shown curve by eye. The straight line of 0 slope fits the data for the ring-absent condition quite well suggesting that a flicker of the second stimulus field without the ring at the various delay intervals has little or no systematic effect upon recognition accuracy of the form presented in F 1. The data for the ring-present condition show a marked masking effect when the ring stimulus is presented after the form at delays under approximately 100 msec. The function appears quite flat from 0 to 50 msec. delay, indicating uniform masking in this range, and it is important to note that the masking effect in this delay range is essentially the same as presenting the form with the ring surrounding it concurrently or on the same stimulus card.

---

[3] These analyses were based on correct recognitions irrespective of *S*'s subjective confidence rating. Separate analyses of the confidence ratings provided no new information over and above that obtained from the correct forced-choice recognitions. The *S*s' confidence in their recognitions was found to describe essentially the same functions.

Discussion

The data are quite clear in showing pronounced retroactive masking effect exerted by the ring stimulus for delay intervals under 50 msec. and further that the masking effect is dissipated rather rapidly somewhere between 50 and 100 msec. delay.

The data obtained from the two groups of *S*s are in close correspondence where the delay intervals were the same. Perhaps the most significant finding with respect to the masking effect is that maximum masking appears to occur when ring and form are presented concurrently. This finding has major implications for the interpretation of the process underlying masking. In the past masking phenomena have been described in terms of erasure by the second stimulus of the first or displacement of the first stimulus by the second in perception. The finding that maximal masking occurs in the concurrent presentation argues against an erasure or displacement hypothesis. Rather, it suggests an interpretation in terms of visual or perceptual persistence and a lag time in the perceptual process itself.

The finding that perception of the letters A, T, or U is impaired when they are presented inside of a ring stimulus is rather easily understood. It suggests that the presentation of one of these letters in a clear white uncluttered field is an easier perceptual task for recognition than the perception of these letters inside of a black ring. Further, the finding that when the ring stimulus is delayed 50 msec. after termination of the letter presentation, recognition impairment is almost as much as when the two stimuli occur concurrently suggests that the perception of the form had not been completed 50 msec. after termination of the stimulation. By the time perception does occur, ring and form appear to be present together requiring the perceptual task of distinguishing the form within the ring. In other words, the basis of the masking effect is that by the time perception occurs ring and stimulus, at short-delay intervals, are present together to form the percept.

Before speculating upon the perceptual-attentional processes that may underlie this phenomena, we deemed it wise to perform another experiment checking upon this interpretation of the basis of masking; namely, the simultaneous presence of the ring and form in the percept. If this interpretation of the impairment in perception is correct, then it would appear that proactive masking should have essentially the same results as retroactive masking. Presenting the ring stimulus prior to presenting the form should describe essentially the same masking function as when the ring follows the form. Either sequence should lead to the percept of the letter surrounded by the ring and thereby constitute a more difficult perceptual task for recognition. The persistence of the time relationships should be essentially the same for either forward or backward masking.

EXPERIMENT II

Method

Subjects.—Ten young adult college students (three males) served as *S*s. Five were assigned to Group C (one male) and the remaining five to Group D. All *S*s were paid a fixed sum for their services and had volunteered for the experiments. They were again given practice sessions on form recognition that did not involve masking arrangements. All had normal or corrected to normal visual acuity.

Procedure.—The procedure, stimulus materials, and apparatus were identical to those employed in Exp. I. The only difference was that under the ring-present conditions in this experiment the ring preceded the letter forms by the various delay intervals. The ring stimulus again was presented for a constant duration of 100 msec. and the duration of the letter exposures was adjusted as before to the individual *S*'s base level. For the 10 *S*s the average base-level duration was 35 msec. with a range of 19 to 55 msec. For the *S*s in Group C the ring occurred concurrently and also preceded the form by lead times of 0, 10, 25, 50, and 250 msec. The corresponding lead times for the *S*s in Group D were 0, 60, 70, 80, and 250 msec.

Results

Separate three-way analyses of variance (ring present vs. absent, delay intervals, and *S*s) were performed for the Group C and Group D data since some delay intervals were not shared in common by these groups. In both groups all three main effects were significant ($p < .01$) as was also the interaction between conditions and delay intervals ($p < .001$). The only significant interaction involving *S*s was the *S*s × Condition interaction in Group C. Inspection of the data revealed that all five *S*s showed an appreciable masking effect at delay intervals under 50 msec. but two of the *S*s showed a much more pronounced masking effect than did the other three.

To facilitate comparison between Exp. I and Exp. II the data from this latter experiment have also been plotted in Fig. 1. Whereas in Exp. I the abscissa represents the delay in milliseconds by which the ring stimulus followed termination of the form, in interpreting the data from Exp. II delay interval represents the time between termination of the ring stimulus and presentation of the form. The dotted lines in Fig. 1 represent the two sigma ranges bounding the designated curves. The sigma value is 3 and was obtained by averaging the error variances for the four groups in the two experiments. Examination of the Exp. II data in Fig. 1 reveals that it corresponds quite closely with the data points from Exp. I. For both the ring-present and the ring-absent conditions, the data points for both experiments seem about equally well fitted by a straight line of 0 slope for the ring-absent condition and by the visually fitted curve for the ring-present condition.

Discussion

The results for both experiments are quite clear in demonstrating three important points. First, both forward and backward masking effects of a ring stimulus upon recognition of simple forms are found when improved indicator methodology is used, methodology that is relatively independent of effects that can be attributed to shifts in *S*'s subjective criterion for report. Second, forward or backward asynchronies of the presentation of the form and of the ring of approximately 50 msec. or less produce an impairment in form recognition that is comparable in magnitude to the impairment obtained when ring and form are presented concurrently. And third, the masking effect of the ring appears to be symmetrical in time in that the forward masking effect is highly similar if not identical to the backward masking effect.

The finding that the maximum forward or backward masking effect is directly comparable to the amount of impairment in form recognition that occurs when ring and form are presented together most strongly suggests that the "masking" is attributable to greater difficulty in recognizing a form surrounded by a black ring than when the form is presented alone in an uncluttered field. In other words an exposure duration that yields 70% form recognition when the form occurs in any empty field is insufficient when the perceptual task requires recognition of the form surrounded by a circle or ring. This latter is a more complex stimulus and as ample evidence has shown, increased exposure durations are required for equal recognition accuracy of more complex stimuli (Hake, 1957).

This explanation of the "masking" effect poses other problems. Specifically, the question is raised as to how a ring stimulus that precedes or follows a form by intervals of 50 msec. or more comes to be perceived as though the two stimuli had been presented together or concurrently?

Three possible mechanisms are suggested. The first is that the visual system is organized so that if a stimulus follows a previous stimulus within a critical lag, the transmission of the second stimulus through the end organs and the visual tracts is facilitated. Thus the second stimulus overtakes the preceding stimulus and they arrive together at the locus or loci in the CNS where perception occurs. We have been unable to locate electrophysiological recording studies that bear directly on this issue, but techniques such as those employed by Lindsley and Emmons (1958) are capable of modification to shed light on this possibility.

A second possibility is that the visual-perceptual system is time gated. That is, visual perception occurs or is organized on the basis of chunks of sensory inputs of between 50 and 100 msec. in size. This time chunking of sensory inputs need not be discrete in that successive 100 msec. of inputs are successively organized. Rather, the process may be one where visual perception is organized on the basis of 100 msec. of input but the successive chunks are greatly overlapping with the preceding and following ones. Similar conceptions of this time-ordering process have been proposed by White and Cheatham (1959) and by Callaway and Alexander (1960) in terms of their "neuronic shutter" theory. (Stroud [1956] has suggested a somewhat similar idea in his quantal theory of consciousness.) In addition to the evidence supplied by these investigators other suggestions that the visual-perceptual system deals or treats stimulation in such time chunks is found in the Bunsen-Roscoe law that holds for durations under 100 msec. and in the recent work on the temporal luminance-summation, contrast-reduction hypothesis (Eriksen & Hoffman, 1963; Eriksen & Steffy, 1964; Thompson, in press). Also Ogle (1963) has shown that little impairment in depth perception occurs when disparity cues to the separate eyes are delayed by as much as 50 msec.

A third possible explanation may be found in attentional processes. Recent research in our laboratory has suggested that part of the time involved in perception can be attributed to an attentional focusing process. Broadbent's (1957) model of the attentional system provides for brief storage of sensory inputs before being attended, and recently Eriksen and Johnson (1964) have shown the feasibility of such a process in auditory perception. Ongoing research in our laboratory suggests that a finite time period is required for the attentional system to focus upon a sensory input and that the time required for focusing depends upon the level of discrimination required. The more complex or difficult the discrimination the longer the time required to focus to the necessary level. While this is highly speculative, the possibility does exist that the effective simultaneity of inputs differing by 50 msec. or more may be due to the focus time of the attentional process. When the first input or stimulation arrives, the attentional process begins to focus upon it, but due to the latency and time involved in the focusing process, the second input or stimulation has arrived before focusing is completed. Thus when the input is attended both stimulations are effectively present together.

The choice between these three possible explanations or maybe others that have not been considered, cannot be made at the present time due to lack of adequate evidence. However, the results of the present experiments have moved the problem of masking to a new level where an explanation must be sought. The present data suggest quite clearly that the basis of masking is due to the more difficult discrimination required by the simultaneous occurrence of the form within a ring. The problem remaining now is to account for the mechanisms by which asynchronies of 50 to 100 msec. are treated by this perceptual system as though they were simultaneous.

Before accepting the above conclusions note must be taken of some apparent conflicts between the data of the present experiments and results reported by other investigators. Averbach and Coriell (1961) presented a display consisting of 16 letters in two rows of eight each. They report that when one of the letters was surrounded by a ring concurrently there was no impairment in the recognition of this particular letter. In fact, only slight impairment in ability to report the encircled letter was noted at 50-msec. delay of the ring but by 100-msec. delay, recognition of the letter was markedly impaired. This result would appear to be in marked conflict with that of the present study but the apparent contradiction can be quite simply resolved. While the duration of exposure of the stimuli in the present experiments was sufficient only for approximately 70% recognition for single-letter exposures, in the Averbach and Coriell study the display was exposed for 50 msec. at a luminance of 70 footlamberts (ftl.), a very strong stimulation and one well above that necessary for 100% correct single-letter identification. It was not of course sufficient for accurate report of all 16 letters. However, if the display contained a circle or ring around one of the letters when it occurred, the effect of the ring would be to direct *S*'s attention to that particular letter rather than the other 15, and since the duration of exposure was more than enough for perfect recognition of single letters, *S* could well report the letter even though the ring may have made it less legible or easily recognized than the other 15. The effect of the ring at 50 and 100 msec. can be further understood in terms of their own data on visual persistence. With such an intense stimulation, well beyond that necessary for single-letter identification, visual persistence is still quite strong at 50 msec. but the interference effect of the ring has become slightly more prominent due to some decay of the stimulus trace. By 100 msec. the decay of the original input has now reached such a level that the occurrence of the ring constitutes effective interference in its recognition. Beyond 130 msec. there is little indication in their data that masking is occurring. There is no evidence presented to show that recognition performance for a ring presented 130 msec. or later differs significantly from the effect of the bar marker they employed in their other experiment that did not produce masking.

Kolers (1962) and Kolers and Rosner (1960) reported an essentially U-shaped masking function where masking of a disk stimulus by a succeeding ring was minimal at a few milliseconds delay, becoming maximal at about 50 msec. and then disappearing in the region of 80 msec. The data from these studies are extremely difficult to compare to the present data or to any of the explanations examined, both due to the great differences in stimulus and task and to the very poor indicator methodology. No attempt was made to assess *S*'s false-alarm rate or to assure that he maintained a given criterion not only within sessions but across sessions in different conditions. As was noted earlier the frequency with which the disk was reported masked varied over 50% for identical stimulus and delay conditions depending solely upon the range of delays an *S* was judging within a session. (This 50% variation attributable to what the authors term anchoring effects is comparable in magnitude to the maximum masking effects obtained in their experiment.) It appears quite likely that *S*'s criterion for reporting the presence of a disk at short delays of the masking ring was quite different than his criterion for reporting the presence of a disk at longer delays. In the absence of data on false-alarm rates it is impossible to interpret either of these experiments in terms of masking phenomena.

REFERENCES

ALPERN, M. Metacontrast: Historical introduction. *Amer. J. Optom.*, 1952, **29**, 631–646.

AVERBACH, E., & CORIELL, A. S. Short-term memory in vision. *Bell Sys. tech. J.*, 1961, **40**, 309–328.

BROADBENT, D. E. A mechanical model for human attention and immediate memory. *Psychol. Rev.*, 1957, **64**, 205–215.

CALLAWAY, E., III, & ALEXANDER, J. D., JR. The temporal coding of sensory data: An investigation of two theories. *J. gen. Psychol.*, 1960, **62**, 293–309.

ERIKSEN, C. W., & COLLINS, J. F. Backward masking in vision. *Psychon. Sci.*, 1964, **1**, 101–102.

ERIKSEN, C. W., & HOFFMAN, M. Form recognition at brief durations as a function of adapting field and interval between stimulations. *J. exp. Psychol.*, 1963, **66**, 485–499.

ERIKSEN, C. W., & JOHNSON, H. J. Storage and decay characteristics of nonattended auditory stimuli. *J. exp. Psychol.*, 1964, **68**, 28–36.

ERIKSEN, C. W., & STEFFY, R. A. Short-term memory and retroactive interference in visual perception. *J. exp. Psychol.*, 1964, **68**, 423–434.

HAKE, H. W. Contributions of psychology to the study of pattern vision. *USAF WADC tech. Rep.*, 1957, **2**, 277–294.

KOLERS, P. A. Intensity and contour effects in visual masking. *Vis. Res.*, 1962, **2**, 277–294.

KOLERS, P. A., & ROSNER, V. S. On visual masking (metacontrast): Dichoptic observation. *Amer. J. Psychol.*, 1960, **73**, 2–21.

LINDSLEY, D. B., & EMMONS, W. H. Perception time and evoked potentials. *Science*, 1958, **127**, 1061.

OGLE, K. N. Stereoscopic depth perception and exposure delay between images to the two eyes. *J. Opt. Soc. Amer.*, 1963, **53**, 1296–1304.

RAAB, D. H. Backward masking. *Psychol. Bull.*, 1963, **60**, 118–129.

STEFFY, R. A., & ERIKSEN, C. W. Short-term, perceptual-recognition memory for tachistoscopically presented nonsense forms. *J. exp. Psychol.*, 1965, **70**, 277–283.

STROUD, J. M. The fine structure of psychological time. In H. Quastler (Ed.), *Information theory in psychology*. Glencoe, Ill.: Free Press, 1956. Pp. 174–207.

SWETS, J. A., TANNER, W. P., & BIRDSALL, T. G. Decision processes in perception. *Psychol. Rev.*, 1961, **68**, 301–340.

THOMPSON, J. H. What happens to the stimulus in backward masking? *J. exp. Psychol.*, 1966, in press.

WHITE, C. T., & CHEATHAM, P. G. Temporal numerosity: IV. A comparison of the major senses. *J. exp. Psychol.*, 1959, **58**, 441–444.

(Received June 18, 1964)

Journal of Experimental Psychology
1965, 70, No. 4, 352–356

COMPETING RESPONSES AND THE PARTIAL-REINFORCEMENT EFFECT[1]

DONALD F. McCOY[2] AND MELVIN H. MARX

University of Missouri

2 groups of 11 rats each were trained and extinguished in a straight runway. Group 50 was reinforced on 50% of the training trials; Group 100 received continuous reinforcement. Competing responses and response latencies were recorded by a photoelectric system. The usual PRE was obtained and competing responses were negatively related to speeds. Weinstock's (1954) hypothesis was not supported, in that Group 50 showed no competing-response habituation on nonreinforced trials.

One of the more influential attempts to account for the partial-reinforcement effect (PRE) in terms of competing responses has been suggested by Weinstock (1954, 1958). According to this notion, conditions of nonreinforcement during training afford the opportunity for competing responses to occur, and thereby to habituate (i.e., drop out). Competing-response habituation is thus seen as the critical factor in the subsequent superior extinction performance of partially reinforced *S*s. Continuously reinforced controls, on the other hand, have no comparable opportunity for the habituation of competing responses in training, so that the onset of extinction favors rapid development of competing responses and consequently diminishing instrumental response strength.

The purpose of the present research was to provide a direct test of the competing-response-habituation hypothesis. Specifically, this experiment replicated the earlier Weinstock study with the addition of (*a*) empirical measurement of competing responses in accordance with a stated definition, and (*b*) analysis of the relationship between the degree of competing-response habituation and subsequent resistance to extinction.

METHOD

Subjects.—The *S*s were 22 naive female albino rats acquired from the Holtzman Company, which were 90 days old at the beginning of the study.

Apparatus.—The apparatus was a straight, enclosed Masonite runway consisting of a 1-ft. start box, an 8-ft. stem, and a 4-ft. goal box. All sections of the runway were 5 in. high and 3 in. wide with the exception of the goal box, which was 5 in. high and 5 in. wide. A manually operated start-box door and an automatically controlled goal-box door separated adjacent runway sections. Both were appropriately closed during the course of a trial to prevent retracing. The start box and stem were covered with removable Masonite tops. The goal box was covered with opaque Plexiglas, behind which were spaced three 6-w. bulbs, 1 ft. apart and 6 in. from each end. This system permitted diffuse goal-box illumination. Along the runway stem were spaced eight photoelectric cells, 1 ft. apart and 6 in. from each end. The goal box contained an additional four photocells, also placed 1 ft. apart and 6 in. from each end. Interruption of each infrared beam activated 1 pen of a 20-pen Esterline-Angus recorder, thereby permitting equally spaced measures of *S*'s activity. Standard Electric timers were used in conjunction with the photocells in the following manner: Running time was measured from the interruption of the first beam in the stem to the interruption of the ninth beam, located 6 in. inside the goal box; goal time was recorded from interruptions between the ninth and twelfth beams, respectively, the latter photocell being located 6 in. from the goal-box terminus. A retractable burrette protruded 1 in. into the end of the goal box and served as a drinking tube through which water reinforcement was delivered. The "off" beam was placed 5 in. from the drinking tube to prevent the development of discriminative goal-box performance on reinforced and nonreinforced trials since the tube was not present during the latter condition.

---

[1] This research was supported by United States Public Health Service Research Grant MH00817 (now HD00895) from the National Institutes of Health and by Research Grant GB229 from the National Science Foundation.

[2] National Institute of Mental Health Predoctoral Fellow.

Procedure.—Eleven *S*s were randomly assigned to the 50% reinforcement group (Group 50) and the remaining 11 to a continuous-reinforcement group (Group 100). Nonreinforced trials for Group 50 were assigned randomly and separately for each *S*. All *S*s were reinforced on the first 3 training trials. Only 1 trial was administered on the first day. Three trials were then given daily for the next 17 days. Training was halted after 52 trials because both groups appeared to have reached asymptote. On reinforced trials the drinking tube was present when *S* entered the goal box and was subsequently available for 20 sec. On nonreinforced trials the tube was absent and *S* was delayed in the goal box for 20 sec. A minimum of 30 min. elapsed between successive trials for each *S*, during which time it was retained in the home cage. According to this schedule, experimental sessions were administered under 22.5–23.0 hr. water deprivation.

Extinction consisted of three daily trials for a total of 7 days. The running order and intertrial interval remained the same as in training. During extinction the drinking tube was completely removed from the experiment. An *S* was withdrawn from the runway on any trial in which it failed to enter the goal box within 90 sec. after the start door opened, and was removed from the experiment following four such successive failures.

Competing responses were scored by means of a plastic template placed over the Esterline-Angus tape. They were defined by any of the following criteria: (*a*) a pause of .5 sec. or more at the location of any photocell; (*b*) a pause of .5 sec. between photocells; (*c*) any reversal of direction as shown by a reversal of the photocell blip sequence.

RESULTS

Acquisition and extinction speeds.—The time measurements were converted to speeds for analytic purposes. An overall analysis of variance of running speeds (RS) in training produced no reliable group differences, but Group 50's performance was coincident with the trend described by Spence (1960); i.e., it showed early suppression and late training facilitation relative to the Group 100 controls. With the appropriate error term from the main analysis, the late training superiority of Group 50 proved to be significant, $F\,(1, 20) = 23.10$, $p < .01$ whereas the early suppression did not. As anticipated, the Groups × Trials interaction was reliable, $F\,(16, 320) = 3.19$, $p < .01$. In the goal speed (GS) measure Group 50 ran consistently, but not reliably, slower than Group 100. The group difference fell just short of the level required for statistical significance, $F\,(1, 20) = 4.18$ (4.35 required) but this trend approximates that reported by Goodrich (1959) and by Wagner (1961).

The superior speeds maintained by Group 50 throughout extinction are concurrent with the usual PRE finding. This well-sustained performance proved to be highly reliable in both RS, $F\,(1, 20) = 29.92$, $p < .01$, and GS, $F\,(1, 20) = 32.68$, $p < .01$. PRE was greater in the goal box and was apparently magnified by the surprisingly limited initial resistance to extinction displayed by Group 100.

Competing responses.—Figures 1 and 2 show the mean number of competing responses during training and extinction in the runway and goal box, respectively. The relative positions of the two groups in terms of competing-response activity can be anticipated from their respective RS and

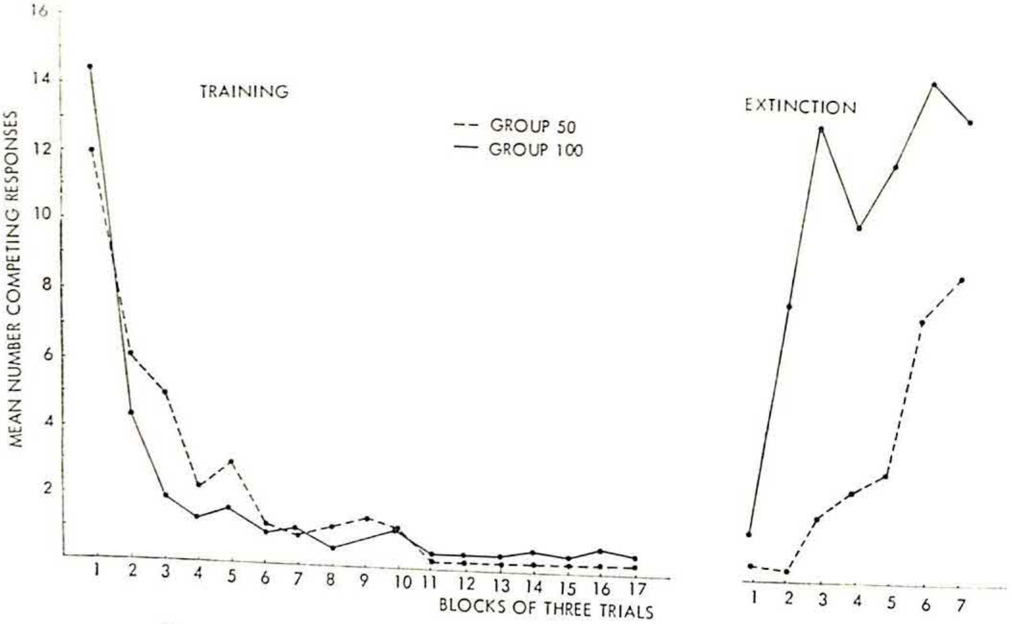


FIG. 1. Mean number of competing responses in the runway.

GS performances. As Fig. 1 indicates, Group 50 made more competing responses in the runway during early training, but consistently fewer by Block 11. This trend corresponds with the early suppression and late facilitation of RS demonstrated by this group. While the overall group differences for this measure were nonsignificant, the Groups × Trials interaction was reliable, $F(16, 320) = 2.24$, $p < .01$. Figure 1 also indicates that the higher RS maintained by Group 50 in extinction was coupled with the execution of fewer competing responses in the runway, $F(1,20) = 8.69$, $p < .01$.

Figure 2 shows a relatively greater number of competing responses for Group 50 over combined reinforced and nonreinforced training trials in the goal box. This difference was

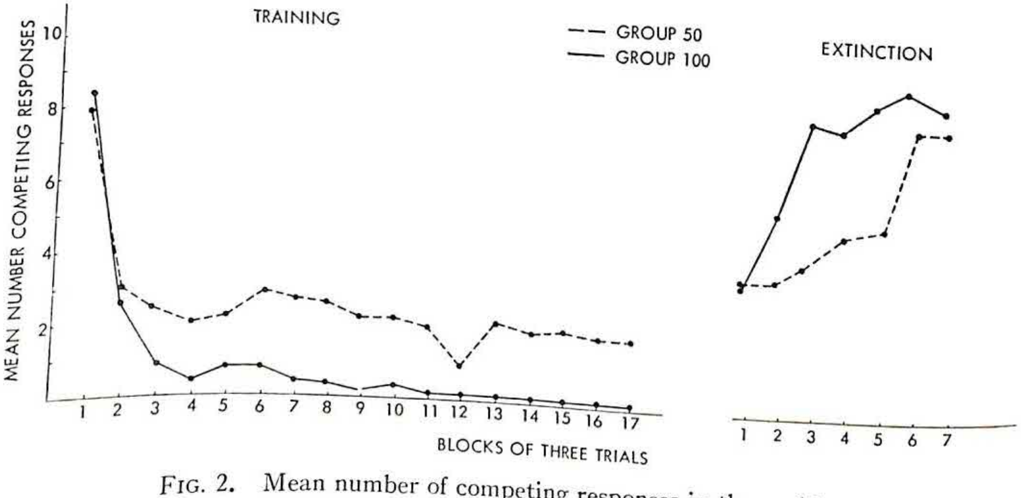


FIG. 2. Mean number of competing responses in the goal box.

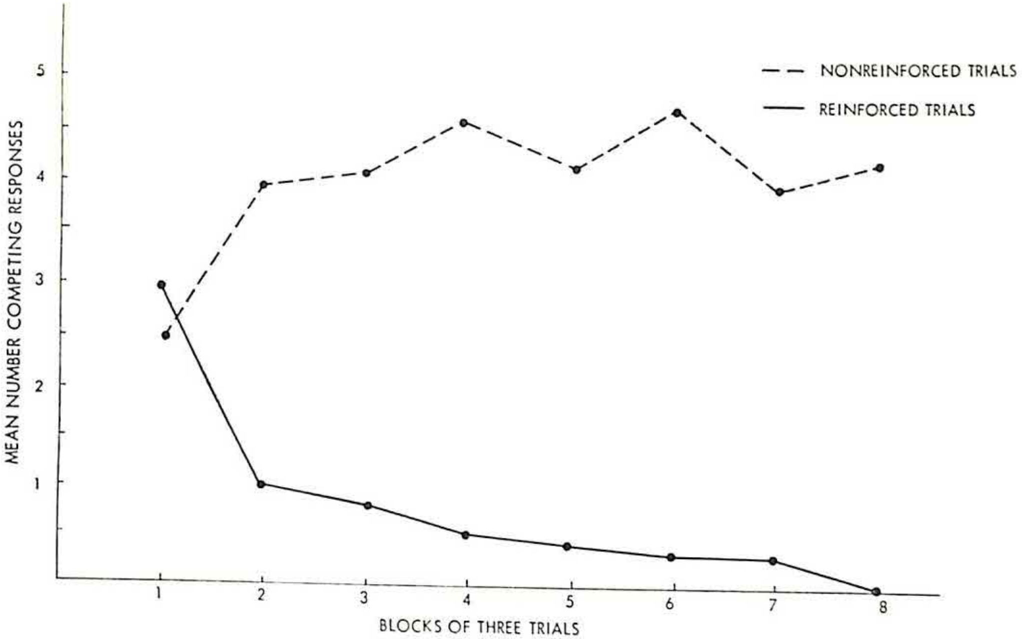


FIG. 3. Competing responses in the goal box for Group 50 on reinforced and nonreinforced trials.

highly significant, F (1, 20) = 41.98, $p < .01$, and is in accordance with the trend toward slower running in the goal box demonstrated by Group 50 in training. Again, the making of significantly fewer competing responses by Group 50 in extinction, F (1, 20) = 7.16, $p < .01$, is consistent with the PRE finding in GS.

Figure 3 is of special interest in that it compares goal-box competing responses for Group 50 on reinforced and nonreinforced training trials. Although the difference between the two conditions was highly reliable, F (1, 20) = 312.35, $p < .01$, of greater interest is the fact that a disproportionate number of competing responses occurred on nonreinforced trials with no evident habituation. The failure of these nonfunctional (competing) goal-box activities to show a habituation effect over nonreinforced training trials has also recently been reported by Longstreth (1964). The reinforced-trials curve reflects a habituationlike effect, but presumably this is due to the increased amount of time spent drinking. Finally, it is important to note that the differences found in Fig. 3 are not attributable to discrimination by *S*s in Group 50 of the presence or absence of the drinking tube on reinforced and nonreinforced trials, respectively. A separate analysis of this group's GS performance on reinforced vs. nonreinforced trials yielded an F of only .13. As might be expected from this value, performance under that condition was virtually equivalent, indicating that the differences found in Fig. 3 are not artifactual.

DISCUSSION

In spite of the consistency of the present training and extinction data with those previously reported (e.g., Lewis, 1960; Spence, 1960), Weinstock's habituation notion receives no support from our data. The finding that Group 50 made more competing responses in the goal box *throughout* training and yet showed the usual PRE is inconsistent with the competing-response-habituation notion. A further incongruency stems from the discovery that competing-response habituation in the goal box was related to conditions of reinforcement rather than to nonreinforcement. Thus if resistance to extinction is assumed to be at all related to the habituation of competing responses in training, it is difficult to explain why Group 100, with twice as many reinforcements and fewer competing responses, was not more resistant to extinction.

It is worth noting that much of the current dissatisfaction with competing-response interpretations (e.g., Marx, 1963) stems from the problems encountered in attempting to define a competing response. The definition employed here served simply as an initial attempt to meet this issue and cannot be considered complete in this sense. A second point concerning competing-response definition questions the plausibility of including activity occurring in an empty goal box as competing activity since it is not at all clear what this behavior competes with. We therefore suggest distinguishing between behavior which directly interferes with an ongoing instrumental response and activity which occurs subsequent to, but does not directly interfere with instrumental behavior. Unfortunately, the manner in which the present data were collected precluded the delineation and exploration of any such relationship. Parenthetically, it is informative to note that Weinstock made no such distinction. Thus, the differential reinforcement and nonreinforcement performances shown in Fig. 3 are directly relevant to his competing-response position.

The present data are consistent with the mediational response hypothesis of Hulse and Stanley (1956) and the frustration hypothesis of Amsel (1958). Hulse and Stanley's assumption that partially reinforced *S*s learn to make "other, mediational responses" under conditions of nonreinforcement receives indirect support, but there is in the present experiment no direct evidence that the partial *S*s actually were doing the "something else" that presumably preserves the instrumental response from extinction. Amsel's assumption of the conditioning of r_f to alley stimuli in training is consistent with the early inferiority and later superiority of the partially reinforced *S*s, as well as their typically superior extinction performance, but again there is no direct evidence of such conditioning.

REFERENCES

AMSEL, A. The role of frustrative nonreward in noncontinuous reward situations. *Psychol. Bull.*, 1958, **55**, 102–119.

GOODRICH, K. P. Performance in different segments of an instrumental response chain as a function of reinforcement schedule. *J. exp. Psychol.*, 1959, **57**, 57–63.

HULSE, S. H., JR., & STANLEY, W. C. Extinction by omission of food as related to partial and secondary reinforcement. *J. exp. Psychol.*, 1956, **52**, 221–227.

LEWIS, D. J. Partial reinforcement: A selective review of the literature since 1950. *Psychol. Bull.*, 1960, **57**, 1–27.

LONGSTRETH, L. E. Partial reinforcement effect and extinction as a function of frustration and interfering responses. *J. exp. Psychol.*, 1964, **67**, 581–586.

MARX, M. H. The need for more specific formulations of the competing response interpretation of extinction. *Psychol. Rep.*, 1963, **12**, 729–730.

SPENCE, K. W. *Behavior theory and learning.* Englewood Cliffs, N. J.: Prentice-Hall, 1960.

WAGNER, A. R. Effects of amount and percentage of reinforcement and number of acquisition trials on conditioning and extinction. *J. exp. Psychol.*, 1961, **62**, 234–242.

WEINSTOCK, S. Resistance to extinction of a running response following partial reinforcement under widely spaced trials. *J. comp. physiol. Psychol.*, 1954, **47**, 318–322.

WEINSTOCK, S. Acquisition and extinction of a partially reinforced running response at a 24-hr. intertrial interval. *J. exp. Psychol.*, 1958, **56**, 151–158.

(Received June 4, 1964)

Journal of Experimental Psychology
1965, 70, No. 4, 357–359

EFFECT OF DELAYED PUNISHMENT ON AN IMMEDIATELY REWARDED RESPONSE IN HUMANS[1]

R. K. BANKS AND M. VOGEL-SPROTT

University of Waterloo

The effect of delayed punishment on the frequency of an immediately rewarded response was explored using 40 human Ss. Punishment delayed by 30, 60, and 120 sec. was found to be less effective than immediate punishment. The results were considered to be consistent with a gradient interpretation of the effect of delayed punishment.

Solomon (1964) recently argued for the existence of a gradient effect of delayed punishment (DP) similar to that of delayed reward. Although the DP gradient may be readily deduced and predicted in theory, empirical support for this phenomenon is very limited and is restricted solely to animal behavior. The effects of DP have been examined for incorrect turns in a maze (Bevan & Dukes, 1955; Warden & Dymond, 1931), avoidance behavior in a shuttle box (Kamin, 1959), and bar pressing (Azrin, 1956; Estes, 1944).

The present study was designed to examine the effect of DP on human behavior in a situation in which a response receiving immediate positive reinforcement is subsequently punished. While this particular paradigm has not been investigated at the animal level, it was selected because it appears to be a common punishment situation for humans: the response is made, positive reinforcement is achieved, and at some time after these events, the punishing consequences of the response occur.

METHOD

Subjects.—Forty volunteers enrolled in undergraduate psychology courses at the University of Waterloo served as *S*s. All *S*s were naive about the purpose of the experiment.

Apparatus.—The apparatus consisted of a signal generator, a three-button response panel, a penny-reward dispenser, and a shock generator with leads to finger electrodes. The pieces of apparatus were interconnected so that one particular series of three presses on the buttons (e.g., 3-1-2) activated the reward dispenser and resulted in immediate delivery of a penny to *S*. The signal generator emitted a mild tone for 1 sec. and was used to signa the onset of each trial. Punishment was administered by the shock generator which delivered a current of 2.25 ma. for .5 sec.

Procedure.—The *S* was seated at a desk in front of the three-button panel. He wa advised that the experiment involved two tests of perceptual-motor skill and that h might at some time receive slight electric shocks from the electrodes placed on the fingers of his left hand.

The first task required *S* to learn the cor rect sequence of three button presses to receive a penny reward. He was told tha each time the tone sounded he could press three buttons in any combination, and that the correct button-pressing sequence would be followed by immediate delivery of a penny reward. This task was considered to be learned when *S* performed the correct response sequence on five consecutive trials. A tria began with the tone and terminated with *S*'s third button press.

When *S*'s performance met the learning criterion, a series of digit-symbol substitution tests, which actually served as a decoy task was presented to *S*. He was instructed to work as accurately and quickly as possible It was suggested by *E* that this task might be somewhat repetitive and that, if *S* wished to dissipate boredom, he might press the button for reward whenever the tone sounded. H was advised to press the buttons the first tim the tone sounded just to "try it out." The instructions in general were designed to focus *S*'s attention on the digit-symbol task and to give him the opportunity to cease button pressing. During the digit-symbol task, *S* was given 15 opportunities (signaled by the tone) to press the buttons. Whenever the response was made it was immediately rewarded and subsequently punished by electric shock. A refusal to respond was counted if *S* failed to press the buttons within 30 sec. after the tone was presented. Thus under punishment training a trial terminated either with the administration of punishment or with a failure to respond within 30 sec.

---

[1] This investigation was supported in part by a research grant from Licensed Beverage Industries Inc. The writers are indebted to Patricia Watson for serving as *E*.

Four DP groups of 0 sec., 30 sec., 60 sec., and 120 sec. were employed. Five intertrial intervals of 15 sec., 45 sec., 90 sec., 150 sec., and 210 sec. were randomly assigned to the series of training and punishment trials and were the same for all *S*s. The relatively long intervals (e.g., 150 and 210 sec.) were used to ensure that punishment in the 120-sec. DP group was not consistently closer to the onset of the next trial than to the response on the previous trial. The experiment was conducted as a series of 10 replications with each replication consisting of one *S* in each of the four DP groups. The *S*s within a replication were allocated by chance to each DP treatment.

RESULTS AND DISCUSSION

The median, mean, and *SD* of the number of punished responses given by the four DP groups are presented in Table 1. Examination of Table 1 shows that the mean and median number of responses increase as the length of DP increases. The difference between the four DP groups was tested by the Kruskal-Wallis one-way analysis of variance of ranks (Siegel, 1956) which yielded an $H\ (3) = 13.76$ ($p < .01$) indicating that delaying punishment has a significant effect on the number of responses given under punishment training. Individual Kruskal-Wallis tests indicated that the 30-sec., 60-sec., and 120-sec. DP groups each differ reliably from the 0-sec. group ($p < .02$) but that they fail to differ reliably from each other ($p > .30$).

TABLE 1

MEDIAN, MEAN, AND *SD* OF RESPONSES DURING 15 PUNISHMENT TRIALS FOR DIFFERENT DELAY-OF-PUNISHMENT INTERVALS

DP Intervals (Sec.)	Mdn.	*M*	*SD*
0	3.0	3.9	3.1
30	12.0	10.1	5.4
60	14.5	10.8	5.4
120	14.7	13.3	4.0

The fact that the 30-, 60-, and 120-sec. DP groups did not differ reliably from each other prevents a definite answer to Solomon's suggestion of a gradient effect of DP. The data are suggestive, however, since the differences between these groups, in terms of both means and medians, are in a direction consistent with a gradient interpretation. Most responses are observed under the 120-sec. delay and least under the 30-sec. delay. Moreover, when the distribution of *S*s displaying the maximum of 15 responses is examined in these three groups, a gradient effect also is observed. Of the 10 *S*s in each DP group, the maximum score was given by four *S*s in the 30-sec. group, five in the 60-sec. group, and six in the 120-sec. group. In addition, the fact that so many *S*s in these groups obtained the maximum score raises the possibility that the failure of the three groups to differ from each other may have been due to a ceiling effect produced by the experimental procedure. If 30 or 40 punishment trials had been employed, the scores could have been more widely distributed. Further research employing more punishment trials, as well as more *S*s per group, would provide an acceptable test of this suggestion.

REFERENCES

AZRIN, N. H. Some effects of two intermittent schedules of immediate and nonimmediate punishment. *J. Psychol.*, 1956, **42**, 8–21.

BEVAN, W., & DUKES, W. F. Effectiveness of delayed punishment on learning performance when preceded by premonitory cues. *Psychol. Rep.*, 1955, **1**, 441–448.

ESTES, W. K. An experimental study of punishment. *Psychol. Monogr.*, 1944, **57**(3, Whole No. 263).

KAMIN, L. J. The delay-of-punishment gradient. *J. comp. physiol. Psychol.*, 1959, **52**, 434–437.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

SOLOMON, R. L. Punishment. *Amer. Psychologist*, 1964, **19**, 239–253.

WARDEN, C. J., & DYMOND, S. A preliminary study of the effect of delayed punishment on learning in the white rat. *J. genet. Psychol.*, 1931, **39**, 455–461.

(Received June 6, 1964)

Journal of Experimental Psychology
1965, 70, No. 4, 360–364

EFFECTS OF CONSUMMATORY RESPONSE PUNISHMENT IN SPATIAL-DISCRIMINATION LEARNING AND RESPONSE FIXATION[1]

CHARLES H. KOSKI AND LEONARD E. ROSS

University of Wisconsin

Rats shocked for consummatory activity (lapping from a drinkometer tube) were compared with nonshock control Ss during spatial-discrimination learning and 4 successive reversals. Shock was individually "tailored" for each shock S according to a behavioral criterion. Reversals for shock Ss were alternately to water with no shock and to the water-shock combination. There was no effect of consummatory-response punishment on original learning. Ss shocked from the beginning of training showed a "fixation" effect in reversing from shock while Ss shocked only after an original-learning criterion had been reached did not. Both early and late shock Ss made more errors than the control Ss in reversing back to the reward-shock combination.

Recently Martin (1963) has reviewed evidence suggesting that a combination of reward and punishment may have facilitating effects per se, in addition to any reinforcing effects that might be present due to *S*'s escape from fear cues in the experimental situation. It is very difficult to investigate the operation of such nonescape facilitation factors in the typical selective learning-punishment situation since shock is usually administered some distance from the goal, a procedure which allows for the simultaneous action of escape-and-approach factors. One way to avoid this problem is to introduce the punishment at the goal by shocking the consummatory response. In a study in which the lapping response was shocked in this manner Martin and Ross (1964) found faster straight-alley running speed, as well as more lapping during the early part of extinction, on the part of shocked animals as compared to noshock control *S*s. It was concluded that the reward-shock combination did have some facilitating effect in the absence of the involvement of escape factors.

The present study extends this consummatory-shock procedure to a T-maze situation in order to investigate the effects of this type of reward-punishment combination on choice and reversal behavior. The performance of *S*s shocked while lapping for water was compared with a reward-only control group during the acquisition and four reversals of a spatial problem. The variable of early vs. late introduction of shock was also investigated. Shock was paired with the water from the beginning of training for half of the shocked *S*s and was introduced only after a learning criterion had been reached for the other half.

METHOD

Subjects.—The *S*s were 36 Sprague-Dawley female rats, 88 days old at the beginning of pretraining.

Apparatus.—The apparatus consisted of animal-holding compartments, start section, T maze, and associated electrical components.

---

[1] This study was supported in part by the Research Committee of the Graduate School from funds supplied by the Wisconsin Alumni Research Foundation, and in part by funds from National Science Foundation Grant GB-765. The research was carried out while the first author was a United States Public Health Predoctoral Fellow.

Components other than the T maze have been previously described (Ross, 1962). The T-maze stem was 20½ in. long and the arms 30 in. in length. The width of the gray stem and arms was 3½ in. and their height 4 in. The T maze was covered with glass and had a hardware cloth bottom except for the brass-rod grid goal areas, which extended 12½ in. from the ends of the arms. Retrace doors were located at the entrance of each arm of the maze and at the entrance to each goal box. A drinking tube mounted in a circular piece of Plexiglas was rotated in and out of each arm by a 100 oz-in., 120 rpm Hurst reversible motor. As the tube turned up and out, the Plexiglas filled the opening. When in the alley the tube orifice protruded ¼ in. from the inset Plexiglas to a point flush with the end of the alley. An electronic drinkometer device described by Koehler and Ross (1963) caused contact with the tube to activate a lap counter, whether or not a shock voltage was present from drinking tube to goal-box floor. The tube automatically withdrew from the arm after 20 shock, or nonshock, laps had been made. A Hunter klockounter recorded time from the first lap contact to the twentieth contact.

Procedure.—The *S*s were placed on a limited water access schedule for 7 days prior to pretraining. Food was always available in the home cage. During pretraining and the experimental period, *S*s were allowed ½ hr. access to water at the end of each session.

On Day 1 of pretraining *S*s had 3 min. of unrestricted drinking in each goal box, the order of placement being random for each *S*. On Day 2 each *S* was given four sessions of 20 laps each, two sessions in each arm, in a counterbalanced order. On Day 3, four trials were given during which *S* was started from its holding compartment and allowed to proceed to one goal area. Both alley doors were open, and a noncorrection procedure was used, but no reward or shock was present. Three groups of 12 *S*s each were formed on the basis of these initial choice trials, matched as closely as possible in terms of their choice behavior.

On the following day training was started. On all trials during this phase, and the following reversal stages of the experiment, *S*s were run with a noncorrection procedure and were removed from the goal box after 20 laps, or at the end of 1 min., whichever came first. The *S*s were given four trials per day for the first 2 days, and eight trials per day during the remainder of the study. Left was correct for all *S*s during original learning.

Shock was introduced for one group, early shock (ES) at the beginning of training. For another group, late shock (LS), shock was introduced after a criterion of 10 consecutive correct choices had been reached. Control group *S*s (C) never received shock. Initial shock level for both the ES and LS groups was initially set at 20 v. since pilot work indicated this level was mildly aversive in terms of disruption of the normally smooth lap rate. Shock was individually "tailored" for each *S*, with increases being made according to an objective criterion of aversiveness. During the experiment the voltage of shock-group *S*s was increased 5 v. for any *S* that completed four consecutive lap trials in less than 6 sec. per trial. Voltage was decreased 5 v. for any *S* that failed to complete 20 laps on any trial, i.e., before 1 min. had elapsed.

Four successive reversals of 32 trials each were carried out following the 13 days of initial discrimination learning. On Days 14–17 (Reversal 1) reward was switched to the right goal and shock turned off for all *S*s. On Days 18–21 (Reversal 2) reward was switched to the left goal box and shock was given to ES and LS *S*s for all 32 trials. The voltages used were the same as those received by each *S* at the end of the original discrimination learning. Reversals 3 and 4 were given in a manner identical to Reversals 1 and 2, with *S*s reversed to the right with water and no shock (Reversal 3) and then back to the left (Reversal 4) where the consummatory responses of *S*s in Groups ES and LS were again shocked as in Reversal 2.

RESULTS AND DISCUSSION

Lap rate.—Lap rate was computed for each *S* on each trial by dividing the number of laps by drinking time, and the means computed from these values for each phase of the experiment are shown in Table 1. The

TABLE 1

LAP RATE MEANS (LAPS/SEC) FOR ORIGINAL LEARNING AND FOUR SUCCESSIVE REVERSAL PERIODS

Groups	Original Learning	Reversals			
		1	2	3	4
C	6.27	7.52	7.11	7.51	7.01
ES	4.51	7.24	2.29	7.48	2.54
LS	4.44	7.05	1.69	7.34	1.51

aversiveness of consummatory response shock during original learning is evident from the fact that the three groups differed significantly, F (2, 33) = 11.33, $p < .005$, with Group C significantly faster than either Group ES or Group LS (Duncan range tests, $p < .05$). The difference between the shock groups did not reach the .05 level.

Analyses of variance of lap-rate performance for each reversal indicated that overall performance was similar ($p > .10$) among the groups during periods where shock was not present, F (2, 33) = 1.48 and .02, for Reversals 1 and 3, respectively. In contrast, overall lap-rate performance differed at the .001 level during reversals when reward and shock were paired, F (2, 33) = 104.71 and 116.14 for Reversals 2 and 4, respectively. Subsequent Duncan range tests showed that Groups ES and LS drank slower than Group C on reversals when shock was present ($p < .05$) and differed from each other at the .05 level only during Shock Reversal 4.

Learning of the original problem.—If the consummatory-response punishment procedures of the present study affected learning of the original problem, such an effect should be seen as a performance difference between Group ES and Groups LS and C in reaching the 10 consecutive correct-response criterion. The means for ES, LS, and C were 27.42, 25.08, and 24.17, respectively, for the trials-to-criterion measure and 5.42, 5.09, and 4.25 for the same groups in the case of errors to criterion. Neither analyses of variance (both Fs < 1) nor Duncan range tests revealed any significant differences.

While there was thus little evidence that the shock procedure influenced acquisition of the original problem to a criterion of 10 correct responses, analysis of group mean errors on postcriterion trials resulted in a significant F at the .025 level—ES, LS, and C group means = 4.58, 6.83, and 2.58, respectively, F (2, 33) = 4.94. Subsequent Duncan range tests showed significance at the .05 level for the LS − C group difference only. Apparently shock did not result in more errors until late in the original period, when shock intensity reached higher levels. (Mean shock voltages for ES and LS on Days 12 and 13 were 26.9 and 25.0 v., respectively.)

Overall, these results indicate that the consummatory-response shock procedures of the present study failed to facilitate acquisition of the problem. If approach facilitation factors were operating, they were canceled by an opposite tendency to avoid the noxious shock.

Fixation effects.—Fixation, as evidenced by increased response persistence following punishment, would appear in the present study as a greater number of errors on Reversals 1 and 3, since errors on these reversals represent responses to the previously correct arm where reward and punishment were paired. Examination of Fig. 1, which presents the mean errors made by each group on each reversal, indicates such a fixation effect for Group ES but not for Group LS. Furthermore, the performance of the LS group appears to be similar to that of the C group on Reversals 1 and 3 (away from shock) and similar to the ES group on Reversals 2 and 4 (back to the reward-shock combination). Statistical analyses confirmed these observations. Separate analyses of variance showed that choice performance differed among the groups during each reversal, $F = 3.79$, $p < .05$; $F = 31.36$, $p < .005$; $F = 7.18$, $p < .005$; and $F = 20.06$, $p < .005$ for Reversals 1–4, respectively, $df = 2,33$ in each

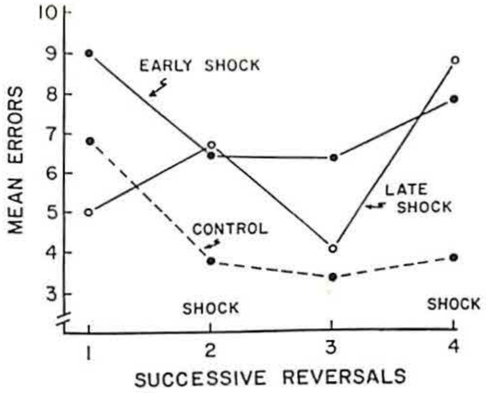


FIG. 1. Mean errors on reversals. (For shock groups reversal was to reward only on Reversals 1 and 3, and to paired reward and shock on Reversals 2 and 4. Control animals never received shock.)

case. Subsequent Duncan range tests indicated that the ES group differed from the C group during all four reversals ($p < .05$) but differed from Group LS only during Reversals 1 and 3, where shock was absent ($p < .05$). Group LS, on the other hand, performed differently from Group C only during Shock Reversals 2 and 4 ($p < .05$).

While each reversal was carried out for 32 trials, it was possible to examine the groups in terms of trials to a criterion of six consecutive correct responses. All *S*s reached this criterion in each reversal with the exception of two ES *S*s and one LS *S*, all in Reversal 4. The pattern of performance with this trials-to-criterion measure was very similar to that shown for errors in Fig. 1, e.g., the means for trials to criterion for the ES, LS, and C groups were 16.92, 12.50, and 13.33 for Reversal 1, and 14.67, 10.50, and 9.58 for Reversal 3. Although the F for Reversal 1 did reach not reach the .05 level, $F\ (2, 33) = 3.12$, $p < .10$, the F on Reversal 3 was significant, $F\ (2, 33) = 6.06$, $p < .025$, and subsequent Duncan range tests showed significant ES − C and ES − LS differences ($p < .01$ and $p < .05$, respectively).

Reversal to shock.—Assuming that the ES differences on Reversals 1 and 3 represent fixation effects from prior punishment, an effect that remains to be considered is the inferior performance of Groups ES and LS relative to Group C on Reversals 2 and 4. The most parsimonious interpretation would seem to be that the poorer performance reflected avoidance of the shock. While the gradual introduction of shock during the learning of a problem did not retard its acquisition (although there are slight, nonsignificant trends in that direction) reversing back to the final intensity after a period without shock might be expected to result in a slower acquisition of the reversal problem. In this respect, it has been demonstrated that the sudden introduction of punishment is more noxious, with correspondingly more disruption of behavior, than a gradual increase to the same final level (Miller, 1960).

If the group differences on Reversals 2 and 4 do indeed reflect avoidance of the shock by the ES and LS groups, it might be predicted that mean trials to the first correct response would be quite similar for all groups on these reversals, i.e., the fact that reversal is to shock should not affect behavior until the first experience with the shock. This proved to be the case. Mean trials to the first reversal trial were practically identical for the shock and control groups on Reversals 2 and 4 (2.08, 2.17, and 1.92; and 2.33, 2.33, and 2.83 for Groups ES, LS, and C, respectively).

The results of the present study indicate the following: Consummatory-response shock, introduced from the beginning of training, does not affect the acquisition of a spatial discrimination in terms of trials or errors to criterion, but does lead to a greater persistence of response, i.e., fixation, during reversal training away from the previously rewarded and punished arm. When shock is introduced after learning, the fixation effects do not appear. Reversal back to the final shock level is slower for both shock groups, presumably due to avoidance of the noxious shock.

Some qualifications of the above should be made. It is quite possible that different methods or rates of increasing shock could facilitate or retard acquisition of the original problem. Also, it should be pointed out that the ES and LS groups received a different number of shock trials during the initial training period, and could have differed in some manner with respect to other aspects of the punishment variable. It would seem somewhat unlikely that the additional ES group shock trials could, by their extra number alone, account for the ES − LS differences, since the additional number is small in relation to the total number of shock trials and the ES − LS difference was found on Reversal 3 after the group's similar Reversal 2 shock experience. Some aspects of the consummatory response (e.g., greater consummatory-response variability during original learning) may be directly related to the fixation effect, but the present study does not allow the identification of such factors.

The fixation effect shown by Group ES cannot be interpreted as a result of the added reinforcing effects of escaping the cues associated with shock since the "end" consummatory response itself was punished. It can be considered in terms of the formulation recently proposed by Martin (1963). Martin suggests that punishment, in the form of fractional anticipatory goal reactions (r_p) and their response-produced stimuli (s_p), may act to (*a*) energize *S*, which could augment a dominant approach response, and/or (*b*) serve as a source of additional cues which become conditioned to the approch response. Thus $r_p - s_p$ conditioning and/or its energizing properties might strengthen approach tendencies and result in a fixation effect. While the present study does not identify which of these factors was operative, or the relative contribution of each if both were effective, it does indicate that these types of punishment approach factors may play an important role in some selective-learning situations.

REFERENCES

Koehler, J. E., & Ross, L. E. A device permitting the use of shock with contact relays. *J. exp. Anal. Behav.*, 1963, **6**, 348.

Martin, B. Reward and punishment associated with the same goal response: A factor in the learning of motives. *Psychol. Bull.*, 1963, **60**, 441–451.

Martin, B., & Ross, L. E. The effects of consummatory response punishment on consummatory and runway behavior. *J. comp. physiol. Psychol.*, 1964, **58**, 243–247.

Miller, N. E. Learning resistance to pain and fear: Effects of overlearning, exposure, and rewarded exposure in context. *J. exp. Psychol.*, 1960, **60**, 137–145.

Ross, L. E. The effect of equal reinforcement of the positive and negative discriminanda of a learned discrimination. *J. comp. physiol. Psychol.*, 1962, **55**, 260–266.

(Received June 9, 1964)

Journal of Experimental Psychology
1965, 70, No. 4, 365–370

SEQUENTIAL DEPENDENCIES IN CHILDREN'S PROBABILITY LEARNING[1]

RICHARD S. BOGARTZ

Institute of Child Behavior and Development, State University of Iowa

2 experiments were conducted to investigate the applicability of a linear operator model to probability learning by preschool children. In Experiment I naive Ss made a motor response under an E-S controlled procedure. 80:20 and 50:50 frequency conditions, with test runs inserted into the event sequences, resulted in positive recency in both groups, but sequential statistics indicated the model was not applicable. In Experiment II 16 Ss, only 4 of whom were naive, were each given a different random 50:50 sequence of colors to predict. Response probabilities conditionalized on the run length during the preceding trials again revealed positive recency. Alternation behavior and position preference found in Experiment I did not occur in Experiment II. A 2-parameter linear learning model gave very good predictions for 18 sequential statistics.

Linear operator models for probability learning (Bush & Mosteller, 1955; Estes, 1959) generally have met with little success when naive *S*s have been used. Sequential predictions have been disconfirmed by response sequences such as negative recency and alternation, as well as by unexpected response-response dependencies (Anderson, 1960; Anderson & Whalen, 1960; Jarvik, 1951; Nicks, 1959). Some *E*s have suggested that these sequential effects may be due to transfer of preexperimentally established verbal habits, possibly due to experience in gambling situations and the like (Estes, 1962). Support for this position is available in that negative recency does tend to drop out over trials (Anderson, 1960; Edwards, 1961). Also, negative recency is not found with extensively practiced *S*s (Friedman, Burke, Cole, Estes, Keller, & Millward, 1963), although in this case Anderson (1964) has offered the alternative explanation that positive recency was being "trained in" rather than that verbal habits producing negative recency were dropping out.

The transfer hypothesis suggests the relevance of studies with children, but although there are a number of child probability-learning studies (McCullers & Stevenson, 1960; Messick & Solley, 1957; Siegel & Andrews, 1962; Stevenson & Weir, 1959; Stevenson & Zigler, 1958), few have studied sequential effects, and those that have have not yielded unambiguous results (Atkinson, Sommer, & Sterman, 1960; Brackbill, Kappy, & Starr, 1962; Kessen & Kessen, 1961; Ross & Levy, 1958). Many of the latter studies have used older children where transfer effects would also be expected. Consequently, the present study was conducted using preschool children as *S*s.

In the two experiments reported here, absence of negative recency was adopted as a minimal prerequisite for additional sequential analysis. Negative recency was not obtained in

---

[1] Portions of this research were conducted during the author's tenure as a National Institute of Mental Health Postdoctoral Research Fellow. Alvin Deibert collected the data in Exp. II. The author is indebted to Norman H. Anderson and Charles C. Spiker for helpful suggestions during the writing of this report.

Exp. I; however, the model failed on other grounds. Inspection of the individual data, a frequently neglected practice, indicated that response-response dependencies such as alternation and position preferences were in part responsible for the observed discrepancies. The individual data obtained using a revised experimental procedure in Exp. II were satisfactory and application of a linear operator model yielded good results.

METHOD

Experiment I

Subjects.—The *S*s were 24 university preschool children. Their ages ranged from 4 yr., 3 mo., to 5 yr., 7 mo.

Apparatus and procedure.—After *E-S* rapport had been established, *S*s were taken individually to a research room and seated in front of a box (17.5 in. wide × 22.25 in. high × 14 in. deep) placed on a low table. A tube, centered in the face of the box 1.25 in. above the surface of the table, permitted delivery of marbles into a container at the front of the box. Two metal response levers (centered between the delivery tube and either side edge of the box), protruded 2.5 in. from the face of the box, and were also 1.25 in. above the table surface. A buzzer and a chime were mounted on the inside rear wall of the box at points equidistant from the two response keys.

The *S*s were instructed to press only one lever as soon as the buzzer sounded. They were told that they would not be correct every time, but that they should attempt to obtain as many marbles as possible by pressing the lever which they thought would release a marble, so that when they had obtained enough marbles they could cash them in for a previously chosen toy.

Two Hunter timers programed the buzzer to sound .3 sec. every 5 sec. On each trial *E* operated a concealed switch to determine which response sounded the chime and released a marble, but on all trials relays prevented rewarding of either correction responses or simultaneous depression of both levers. Responses and rewards were recorded on a Brush event recorder.

Design.—The *S*s were randomly assigned in equal numbers to an 80:20 (Group 80) and a 50:50 (Group 50) frequency condition. All *S*s in Group 80 received the same random 80:20 sequence in Stage 1 (Trials 1–30). Three additional 60-trial blocks constituted Stages 2, 3, and 4. Each was constructed by randomly assigning three test runs to three nonoverlapping and nonadjacent 8-trial blocks within the 60-trial block. A test run consisted of a reward for the less frequently rewarded response, followed by seven consecutive rewards for the other response. Rewards on the remaining 36 trials within each block were then randomly determined so that the reward ratio over the 60 trials was 80:20.

A 3 × 3 Latin-square design permitted three different orderings of the 60-trial blocks.

The trial-sequence construction for Group 50 was similar, except that a test run consisted of a reward for one response followed by four consecutive rewards for the other, leaving 45 trials to be randomly determined within each 60-trial block.

Experiment II

Subjects.—Sixteen university preschool children served as *S*s. Twelve of these served in Exp. I, but Exp. II, which began 5 mo. after Exp. I, employed a change in apparatus, *E*, and experimental room. In the interval, *S*s served in other unrelated experiments.

Apparatus and procedure.—The marble-dispensing mechanism of Exp. I was concealed by a 48 in. wide × 38 in. high brown Masonite board placed on a low table. A microswitch, which served as the response manipulandum, was located at the bottom center of the board. Closing this switch released a marble into a receptacle beneath the switch. On each trial *S* removed the marble to a separate container.

The *S* was seated in front of the apparatus and told that he could win a toy by correctly guessing the color of marbles. A black and a white marble were presented simultaneously. and *S* was required to identify their colors, He was then instructed to guess the color of the next marble after the buzzer sounded but before he pressed the button. Five practice trials with red marbles followed. After practicing, *S* was again shown the two marbles and reminded that "there will only be these kinds of marbles."

The *E* sat behind and to the side of *S* and recorded both responses and marble outcomes, since each *S* received a different random 50:50 sequence of marbles. The 150 trials were paced as in Exp. I.

RESULTS

Experiment I

For half the *S*s in Group 80 the more frequently rewarded response (A_1)

was pressing the right-hand lever; for the other half, the left. Since the frequencies of reward are equal in Group 50, the right-hand response was designed as A_1 for half the *S*s and as A_2 for the others.[2] The reinforcement procedure is *E-S* controlled (Bush & Mosteller, 1955). Let reward mean release of a marble and sounding of the chime. Then, four reinforcing events, E_{11}, E_{21}, E_{12}, and E_{22}, correspond to the following four conjunctions of responses and outcomes, respectively: A_1 and reward; A_1 and no reward; A_2 and reward; A_2 and no reward. The Bush-Mosteller linear model for *E-S* controlled events is $p_{n+1} = \alpha_{fi}p_n + (1 - \alpha_{fi})\lambda_{fi}$, where λ_{fi} is the limit point for Event E_{fi}, and α_{fi} is the rate of learning associated with E_{fi}. It is reasonable to assume that the limit points in this experiment are 1 and 0 so that $\lambda_{11} = \lambda_{22} = 1$ and $\lambda_{12} = \lambda_{21} = 0$. It also seems reasonable to assume that the learning rate for A_1 and reward is the same as that for A_2 and reward, and that the learning rate under nonreward also does not depend on the response, so that $\alpha_{11} = \alpha_{12} = \alpha$ and $\alpha_{22} = \alpha_{21} = \alpha'$.

A check for negative recency can be made by fractionating the data from the test runs to obtain the proportion of *S*s making the A_1 response on Test Trial $N + 1$, given the occurrence of A_1 and reward on all previous N test trials. Thus for Group 80 N ranges from 1 to 7; for Group 50, from 1 to 4. Negative recency would be indicated by a rise and then fall of the obtained values. For Group 80 the values were .78, .95, .93, .96, 1.00, .96, and .98 for $N = 1$–7, respectively. The values for $N = 1$–4 for Group 50 were .47, .86, .83, .90. Thus negative recency was not obtained.

A test of the applicability of the model has been derived which does not require parameter estimation (Anderson, 1959). Under the above assumptions concerning the parameters, if the model holds then

$$p_{1,11}/\Pi_1 + p_{1,21}/\Pi_2 = V_1,$$

where $p_{1,11}$ is the joint probability of A_1 on Trial n, A_1 on Trial $n - 1$, and reward on Trial $n - 1$; $p_{1,21}$ is the joint probability of A_1 on Trial n, A_2 on Trial $n - 1$, and reward on Trial $n - 1$; Π_i is the probability of reward following A_i, $i = 1, 2$; and V_1 is the mean probability of an A_1 response. For Group 80 the estimate of V_1 was .810 and the sequential statistic was 1.299; for Group 50 the corresponding values were .440 and .622. For both groups the gross discrepancy between the two statistics indicates the inapplicability of the model to these data. Inspection of six sequential statistics for each *S* revealed strong alternation behavior or position preference for several *S*s in both groups. These idiosyncratic response patterns were assumed to underlie, at least in part, the failure of the model and Exp. II was designed in an attempt to decrease the likelihood of occurrence of such response-response dependencies.

Experiment II

Since color preferences may occur in probability learning (Nicks, 1959), it is appropriate to preserve information regarding such tendencies. Accordingly, a black prediction is denoted by A_1 and a black marble outcome by E_1. Figure 1 is a plot of the A_1 frequency as a function of the length of the run of preceding E_1's or E_2's, with the number of cases at each point indicated across the top. A clear picture of positive recency is shown, in agree-study by Craig and Meyers (1963), a motor response was used and alternation behavior was observed. When a verbal response was used in Exp. II, and in a subsequent experiment with sequences of colored cards, alternation behavior did not occur, in agreement with the results of Kessen and Kessen (1961). Two other important differences between the two experiments, the difference in the reinforcement procedure and the fact that most *S*s in Exp. II had served in Exp. I, also require additional investigation. This could begin with independent replication of the two experiments.

---

[2] The proportion of response omissions was about .006. These responses were filled in randomly.

It is not clear yet at what age *S*s begin to exhibit negative recency. Craig and Meyers (1963) found no negative recency with kindergarten children; however, many of them were exhibiting alternation with both 60:40 and 80:20 sequences. Their fourth-grade *S*s show negative recency with 80:20 but tend to alternate with 60:40. Older *S*s show even more pronounced negative recency. Subject to additional results from *S*s in the first few school grades, it can be tentatively concluded that *S*s do not exhibit negative recency earlier than the first grade.

REFERENCES

ANDERSON, N. H. An analysis of sequential dependencies. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory.* Stanford: Stanford Univer. Press, 1959. Pp. 248–264.

ANDERSON, N. H. Effect of first-order conditional probability in a two-choice learning situation. *J. exp. Psychol.*, 1960, **59,** 73–93.

ANDERSON, N. H. Comments on Professor Estes' paper. In A. W. Melton (Ed.), *Categories of human learning.* New York: Academic Press, 1964. Pp. 129–144.

ANDERSON, N. H., & WHALEN, R. E. Likelihood judgments and sequential effects in a two-choice probability learning situation. *J. exp. Psychol.*, 1960, **60,** 111–120.

ATKINSON, R. C., SOMMER, G., & STERMAN, M. B. Decision making by children as a function of amount of reinforcement. *Psychol. Rep.*, 1960, **6,** 229–306.

BRACKBILL, V., KAPPY, M. S., & STARR, R. H. Magnitude of reward and probability learning. *J. exp. Psychol.*, 1962, **63,** 32–35.

BUSH, R. R. The application of learning models to interactive behavior. In J. Criswell, H. Solomon, & P. Suppes, *Mathematical methods in small group processes.* Stanford: Stanford Univer. Press, 1962. Pp. 69–73.

BUSH, R. R., & MOSTELLER, F. Stochastic models for learning. New York: Wiley, 1955.

CRAIG, G. J., & MEYERS, J. L. A developmental study of sequential two-choice decision making. *Child Develpm.*, 1963, **34,** 483–493.

EDWARDS, W. Probability learning in 1,000 trials. *J. exp. Psychol.*, 1961, **62,** 381–390.

ESTES, W. K. The statistical approach to learning theory. In S. Koch (Ed.), *Psychology: A study of a science.* New York: McGraw-Hill, 1959, Vol. 2. Pp. 380–491.

ESTES, W. K. Learning theory. *Annu. Rev. Psychol.*, 1962, **13,** 107–144.

FRIEDMAN, M. P., BURKE, C. J., COLE, M., ESTES, W. K., KELLER, L., & MILLWARD, R. B. Extended training in a noncontingent two-choice situation with shifting reinforcement probabilities. In R. C. Atkinson (Ed.), *Studies in mathematical psychology.* Stanford: Stanford Univer. Press, 1963. Pp. 250–291.

JARVIK, M. E. Probability learning and a negative recency effect in the serial anticipa- of alternative symbols. *J. exp. Psychol.*, 1951, **41,** 291–297.

KESSEN, W., & KESSEN, M. L. Behavior of young children in a two-choice guessing problem. *Child Devlepm.*, 1961, **32,** 779–788.

MCCULLERS, J. C., & STEVENSON, H. W. Effects of verbal reinforcement in a probability learning situation. *Psychol. Rep.*, 1960, **7,** 439–445.

MESSICK, S. J., & SOLLEY, C. M. Probability learning in children: Some exploratory studies. *J. genet. Psychol.*, 1957, **90,** 23–32.

NICKS, D. C. Prediction of sequential two-choice decisions from event runs. *J. exp. Psychol.*, 1959, **57,** 105–114.

ROSS, B. M., & LEVY, N. Patterned prediction of chance events by children and adults. *Psychol. Rep.*, 1958, **4,** 87–124.

SIEGEL, S., & ANDREWS, J. M. Magnitude of reinforcement and choice behavior in children. *J. exp. Psychol.*, 1962, **63,** 337–341.

STEVENSON, H. W., & WEIR, M. W. Variables affecting children's performance in a probability learning task. *J. exp. Psychol.*, 1959, **57,** 403–412.

STEVENSON, H. W., & ZIGLER, E. Probability learning in children. *J. exp. Psychol.*, 1958, **56,** 185–192.

(Received June 12, 1964)

Journal of Experimental Psychology
1965, 70, No. 4, 371–378

TWO STAGES OF PAIRED-ASSOCIATE LEARNING AS A FUNCTION OF INTRALIST-RESPONSE SIMILARITY (IRS) AND RESPONSE MEANINGFULNESS (M)[1]

JOHN JUNG[2]

Long Beach State College

Separate studies show that high intralist-response similarity (IRS) aids the response-learning stage (Underwood, Runquist, & Schulz, 1959) and hinders the associative-learning stage (Horowitz, 1962) of paired-associate learning. Neither study examined both stages simultaneously as does the present study. In addition, effects of response meaningfulness (M) on each stage are assessed. 2 levels of each variable were studied in Experiment I. Response learning was superior to associative learning on early trials but only with high IRS; little difference between the 2 types of learning resulted with low IRS. Experiment II replicated only the high-M conditions of Experiment I using separate rather than the same groups of Ss for each type of learning. Results obtained with the 2 methods were highly similar.

Two separate studies seeking evidence for the two-stage analysis of paired-associate learning proposed by Underwood, Runquist, and Schultz (1959) have examined the effects of intralist response similarity (IRS) on the first or response-learning stage (Underwood et al., 1959) and on the second or associative-learning stage (Horowitz, 1962). Measuring Stage 1 by a response recall test after each trial, Underwood et al. found as predicted that response learning is facilitated by high IRS. Noting that paired-associate anticipation learning is generally better with low IRS, these *E*s inferred that while high IRS aids Stage 1, it probably impedes Stage 2.

The Horowitz (1962) study was a direct test of the effects of IRS on associative-stage learning. After each study trial, *S* was provided with a set of cards each containing one of the list responses and instructed to match them on a formboard which contained all of the list stimuli. The results confirmed the prediction that Stage 2 is hindered by high IRS.

The present study combines the Underwood et al. (1959) and Horowitz (1962) studies in order to examine the occurrence of the two stages *simultaneously* as well as to replicate each of the two studies. The examination of the relationship of the two stages to each other is of especial interest since it may shed evidence on which stage occurs first. The conventional paired-associate method involving anticipation of the response to each stimulus as it is presented tends to suggest that the response-learning stage starts first temporally and is completed before or no later than the associative-learning stage. That is, under the anticipation procedure, it is possible to observe the completion of Stage 2 only after (or at the same time) Stage 1 has been completed. It is possible that under some conditions at least, associative learning might actually exceed response learning, especially on early trials. The associative-matching method of Horowitz (1962), unlike the anticipation method, does not "force" the associative stage to be second. By using both the response-recall and associative-matching tests after each trial, it will be possible to examine the relationship between the two stages at successive points in learning.

---

[1] The author wishes to thank Valerie Truscott for conducting Exp. I and assisting in preliminary data analysis.

Portions of this paper were presented at the American Psychological Association Meetings, September 1964, in Los Angeles, California.

[2] Now at York University, Toronto.

The prediction that high IRS would improve response recall (Underwood et al., 1959) was based on considerations derived from studies involving words, e.g., the clustering effect in recall found by Bousfield and Cohen (1955). Similarity in those experiments is one of meaningful similarity among the words; however, if the responses are nonsense syllables, similarity is of a formal nature defined by the extent to which the items within the list employ similar letters. The present study, unlike that of Underwood et al. (1959), utilizes nonsense-syllable responses and it is possible that the mechanisms which affect response recall with meaningful IRS may be absent, minimized, or different under formal IRS. While this difference may be important and should possibly be eliminated, one advantage accruing to the use of poorly integrated units such as nonsense-syllable responses is that they allow more response learning to occur during the experiment than would well-integrated units such as words.

The effects of a second variable, response M, on the two stages will also be examined. Response M has been firmly established as a powerful variable on the response-learning stage but no experimental evidence regarding its effects on the associative-learning stage was presented in the monograph by Underwood and Schulz (1960) which assessed the effects of M in verbal learning. The present study provides evidence of the effects of response M not only on Stage 1 by the response-recall tests but also allows examination of its effects on Stage 2 with the associative-matching tests. Underwood and Schulz (1960) hypothesized that associative probability may be a mechanism by which M could affect the associative stage. Associations from both the stimulus and response elements of a pair make possible the associative learning of that pair. If this process is tenable, the higher the response M (stimulus M as well), the faster the completion of the associative stage, per se.

EXPERIMENT I

Method

Subjects.—Sixty volunteers from introductory psychology at Long Beach State College were assigned in an abcd order to the four conditions. None of the *S*s had served in any previous verbal-learning experiment.

Design.—Four conditions were obtained by combining two levels of the two independent variables, IRS and response M. One group of 15 *S*s served in each of the conditions, High IRS-High M; High IRS-Low M; Low IRS-High M; Low IRS-Low M.

Materials.—Four lists of eight pairs of digit stimuli (1–8) and nonsense-syllable responses from Underwood and Richardson (1956) were used. The lists varied in formal IRS (high and low) and response M (high and low). These lists and details of these compositions are not repeated here (see Underwood & Richardson, 1956). While lower in M than the adjective responses of Underwood et al. (1959), these pronunciable CVC units probably are higher in M than the CCC trigram responses used by Horowitz (1962).

Procedure.—An *S* was instructed to study the digit-nonsense syllable pairs on each of the eight cards which would be presented. No anticipation learning was required although the stimulus was first presented alone for 2 sec., then the stimulus-response pairing was shown for 2 sec. before proceeding to the next paired associate. The order of the pairs varied from trial to trial.

After the complete list had been presented on each trial, *S* was presented with a slip of paper and given a maximum of 45 sec. in which to write down as many responses as he could recall. If he finished within 45 sec., the test was ended for that trial in order to minimize rehearsal. Then, eight cards containing only the list stimuli were arranged in a column on a table in front of *S*. The serial order varied from trial to trial. A deck of eight cards containing only the list responses in a shuffled order was given to *S* who had a maximum of 45 sec. to match these responses with the correct stimuli. He was cautioned against blind guessing and not required to match all eight cards if he lacked confidence in any of the judgments. As soon as *S* finished or 45 sec. had elapsed (this amount of time proved adequate for all *S*s), *E* recorded his matchings while *S* performed a digit-cancellation task. Approximately 30 sec. later the next study trial was given followed by the two tests. This procedure continued for five alternate study and test trials.

Results and Discussion

The mean number of correctly recalled responses and correct matchings for each trial is presented in Fig. 1 and 2 for the four conditions.

The effects of IRS and response M on the two stages were analyzed by two different methods. The first analysis which follows below involves an examination of the effects of the two independent variables on each response measure separately. This analysis provides a direct comparison with Underwood et al. (1959) for Stage 1 and Horowitz (1962) for Stage 2 in regard to IRS. It also permits assessment of the effects of response M on each stage considered alone.

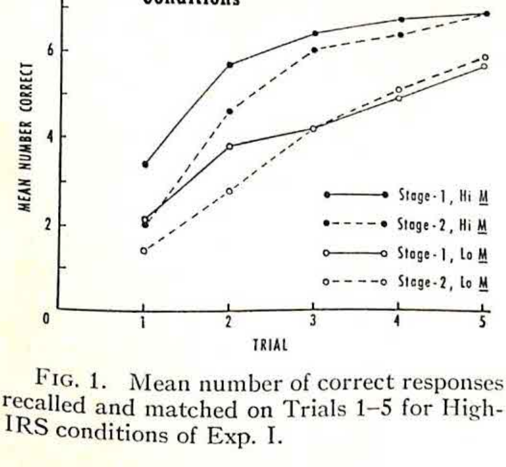


FIG. 1. Mean number of correct responses recalled and matched on Trials 1–5 for High-IRS conditions of Exp. I.

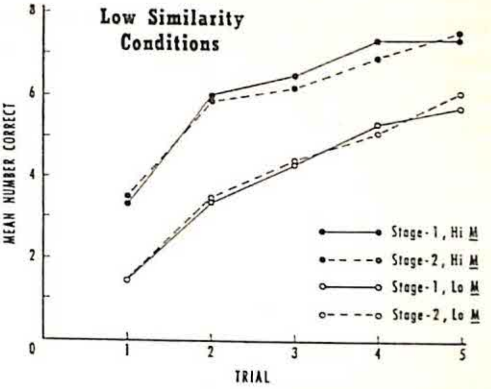


FIG. 2. Mean number of correct responses recalled and matched on Trials 1–5 for Low-IRS conditions of Exp. I.

Response-learning stage.—A 2 × 2 analysis of variance was performed on the response-recall scores on Trials 1–5 for the four conditions. IRS was not a significant variable ($F < 1.0$). The mean number of responses recalled for high IRS (24.9) was slightly less than that for low IRS (25.2) rather than greater. Thus the finding of Underwood et al. (1959) that high meaningful IRS leads to higher response recall is not confirmed with high formal IRS. This is not surprising since IRS is defined quite differently with words and nonsense syllables.

Response M, on the other hand, had a significant influence on response recall. A mean of 29.8 responses was obtained under high M and of 20.3 with low M on Trials 1–5. This difference gave a significant F (1, 56) = 33.30, $p < .01$. These findings agree with those of Underwood and Schulz (1960) concerning the effects of response M on Stage 1. The M × IRS interaction F (1, 56) < 1.0 was not significant.

Associative-learning stage.—A 2 × 2

least no later than associative learning for paired-associate tasks in general.

Some evidence relevant to this conclusion is available from an analysis of the interdependence of the two stages in the present study. A determination was made of the frequencies of each of the following joint outcomes on the two tests used on each trial: Correct Recall Only, Correct Matching Only, Correct Recall and Matching. The proportion of the total number of correct items contained in each of these categories on each trial is presented in Table 1 for each of the four conditions.

An examination of the proportion correct on both recall and matching on Trial 1 shows that only from .18 to .54 of the total items correct, depending on the IRS-M condition, were correct on both tests. Thus, the majority of the items correct on Trial 1 were either correct on recall only or matching only, the combined proportions ranging from .82 to .46. This suggests that on Trial 1 the two stages may be independent processes; an item may be correctly recalled but not necessarily matched and vice versa.

However, by Trial 5 the proportions correct on both recall and matching increase to between .67 and .88, reflecting the fact that learning has occurred. The proportions of the combined correct on recall only and matching only items is reduced to between .33 and .12; thus, the two stages are highly interdependent processes late in learning. The extent of this relationship depends on the particular IRS-M condition, as was the case on Trial 1 when the stages were relatively independent processes.

Although the preceding analysis lends support to a view that degree of training and the stimulus conditions determine the level of interdependence of the two stages, it does not suggest which stage occurs first under those conditions in which the stages appear to be independent processes. A comparison of the proportions of items correct on recall only with those correct on matching only is related to this question. If these two proportions are equal this would suggest that the two stages develop at the same rate; however, to the extent that the proportion of one category exceeds that of the other, the stage represented by the larger proportion may be said to occur first or at least at a faster rate.

Under High-IRS conditions, the proportion of correct recall only exceeds that of correct matching only especially on Trial 1 (High M, .50 vs. .17; Low M, .47 vs. .23). As training continues, however, the proportions approach equality. For the Low-IRS conditions on Trial 1 as well as on later trials, the proportions of correct recall only and correct matching only are about equal and in a few instances matching slightly exceeds recall (High M, .19 vs. .26; Low M, .41 vs. .41). Thus, it appears that IRS level affects the relative rates of development of the two stages on early trials. With high IRS, Stage 1 seems to develop faster than Stage 2; furthermore, the superiority of Stage 1 exists at both levels of M. On the other hand, with low IRS, the rate of development of the two stages is essentially equal for both levels of M with Stage 2 actually surpassing Stage 1 on some trials. Thus, the analyses based on individual item outcomes agree highly with those based on the lists as a whole; both suggest that Stage 1 develops more rapidly than Stage 2 under high IRS but that with low IRS the two stages develop at equal rates.

EXPERIMENT II

It may be objected that the exact relationship of the two stages to each other may be distorted by the procedure employed in Exp. I which required each *S* to undergo both types of tests after each trial. For example, some response learning certainly may occur during the associative-matching test which will serve to inflate the response-recall test score of the next trial. Similarly, the execution of the response-recall test may affect *S*'s performance on the associative-matching test which follows it. A purer experiment utilizing separate groups for each type of test as well as separate groups after different amounts of training for the four stimulus conditions would have required 40 experimental conditions. This more precise design was avoided not merely because it was infeasible but for the reason that the goal of the study was to examine the general, not the exact, relationship of the two stages. Thus, although the dual testing employed in Exp. I might distort the relationship between the stages, it was still possible to determine how the various combinations of IRS-M affected the relationship between the two stages. Furthermore, since one goal was to study the interrelationship of the two stages within an individual *S*, it was mandatory to administer both tests to each *S*.

However, to meet the possible objection to the dual testing of each *S*, a second experiment was performed in which only one type of test was given to each *S*. However, *S*s were still tested after each of the five trials given. Repeated testing of *S* after each of several trials on only one type of test does not seem to be as serious a problem as that of testing an *S* on both types of tests. A replication was made only of the High-M conditions using separate groups of *S*s for recall and matching tests.

Method

Subjects.—Each of the four conditions contained 20 volunteers from introductory psychology classes.

Design.—Half the *S*s received the Low- and half received the High-IRS list (only the High-M lists were used in Exp. II). At each IRS level, half the *S*s were required to free recall the responses and the other half were required to match the stimuli and responses of the list which were provided to them on test sheets.

Materials.—The High- and Low-IRS lists of high M from Exp. I were placed on acetate slides by using commercially produced cutout letters.

Procedure.—All *S*s were instructed that a list of eight number-trigram pairs would be projected on the screen. Prior to the first trial, each condition received explicit instruction as to the nature of their task. Response-recall groups were instructed to learn only the trigrams and not the number-trigram pairings. Associative-matching groups were instructed to learn the specific number-trigram pairings presented. Insofar as *S*s in the present experiment had to attend only to one rather than both types of tasks, the performance on the two tests may differ from that obtained in Exp. I where *S*s were given both tests and thus could not ignore one task while concentrating exclusively on the remaining task.

Response-recall condition *S*s were provided with five slips of blank paper, one for each test trial, on which to record their responses; matching-condition *S*s received five test sheets on which the list stimuli were reproduced along the left-hand margin and the list responses were presented in a row along the top. Different orders of the items were used on each of the five test sheets and none of the stimuli and responses of an actual list pair were positioned on any test sheet in the same ordinal position.

The order in which the list was presented was different for each of the five study trials. Rate of presentation was about 2 sec. per item. The test trial of 45 sec. for either type of test served as the intertrial interval. A total of five alternate study and test trials was given under all conditions.

The administration of each of the four conditions was executed in two separate group sessions in which half the *S*s for a given condition were tested. Orders of presentation for a given list were varied for the two sessions.

Results and Discussion

The mean number correct for recall and matching under low and high IRS on each of the five trials is presented in Fig. 3. In order to compare the

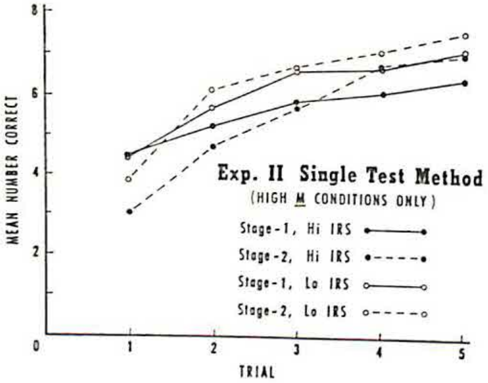


FIG. 3. Mean number of correct responses recalled and matched on Trials 1–5 for High- and Low-IRS conditions of Exp. II.

single and double methods of testing, these results should be directly compared with the two uppermost curves (recall and matching) of Fig. 1 (high IRS) and of Fig. 2 (low IRS). There is high agreement between the results of Exp. I and II although they utilize different testing methods. Both studies show higher performance with low than with high IRS for both types of tests. For response recall, however, the effects of IRS were not significant in either study; Exp. II, $t\ (38) = 1.52$, $p > .10$. It will be recalled that Underwood et al. (1959) obtained significantly higher recall with high IRS but they used words as responses.

Associative-matching results in both Exp. I and II agree with Horowitz' (1962) finding that performance is higher with low IRS. This difference was significant only in Exp. II; $t\ (38) = 1.88$, $p < .05$.

The relationship between the two stages for a given IRS level is also quite similar in both experiments, in spite of the fact that measures of the stages come from the same *S* in Exp. I but from different *S*s in Exp. II. In the Low IRS condition, both experiments show little difference between recall and matching performance; Exp. II, $t(38) = .60$. Under high IRS in Exp. I, recall was consistently higher than matching on early trials but by Trial 5 performance on the two tests was equal. The results of Exp. II on early trials agree with those just described from Exp. I but on later trials matching not only equaled but surpassed recall. Although the overall difference between recall and matching scores for high IRS in Exp. II was not significant, $F\ (1, 38) < 1.0$, the interaction between Type of Test and Trials was significant with $F\ (4, 152) = 6.01$, $p < .01$. Thus, both experiments show recall to exceed matching with high IRS on early trials; neither study shows differences at any point between the two stages under low IRS. This basic relationship is obtained although the methods of assessing the two stages, as well as other experimental procedures, differed in the two studies.

REFERENCES

BOUSFIELD, W. A., & COHEN, B. H. The occurrence of clustering in the recall of randomly arranged words of different frequencies of usage. *J. gen. Psychol.*, 1955, **52**, 83–95.

HOROWITZ, L. M. Associative matching and intralist similarity. *Psychol. Rep.*, 1962, **10**, 751–757.

HOROWITZ, L. M., & LARSEN, S. R. Response interference in paired-associate learning. *J. exp. Psychol.*, 1963, **65**, 225–232.

UNDERWOOD, B. J., & RICHARDSON, J. The influence of meaningfulness, intralist similarity, and serial position in retention. *J. exp. Psychol.*, 1956, **52**, 119–126.

UNDERWOOD, B. J., RUNQUIST, W. A., & SCHULZ, R. W. Response learning in paired-associate lists as a function of intralist similarity. *J. exp. Psychol.*, 1959, **58**, 70–78.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning.* Chicago: Lippincott, 1960.

(Received June 12, 1964)

Journal of Experimental Psychology
1965, 70, No. 4, 379–385

EFFECT OF PERCEPTUAL PRETRAINING ON REVERSAL AND NONREVERSAL SHIFTS[1]

LOUISE SHERLOCK TIGHE[2]

Cornell University

Experimental Ss (1st-grade children) received in succession (a) nonreinforced pretraining in making same-different judgments to stimuli varying in height and brightness, (b) a reinforced-discrimination task in which stimuli varied simultaneously in height and brightness, and (c) either a reversal shift of the initial discrimination ($\frac{1}{2}$ Ss) or a nonreversal shift ($\frac{1}{2}$ Ss). Control Ss received the same discrimination task and shifts but preceded by a picture-completion and a picture-arrangement test. Ss in the experimental-reversal group required significantly fewer trials to reach criterion than Ss in the experimental-nonreversal and control-reversal groups. The control-reversal and control-nonreversal groups did not differ. The results are interpreted within a differentiation theory of discrimination learning.

Adult humans find a reversal shift (RS), which requires response to a previously relevant dimension, easier than a nonreversal shift (NRS), which requires response to a previously irrelevant dimension (e.g., Buss, 1953; Harrow & Friedman, 1958). On the other hand, rats (Kelleher, 1956), chickens (Brookshire, Warren, & Ball, 1961), and rhesus monkeys (Tighe, 1964) have been found to execute an NRS more rapidly than an RS. Studies at the human level indicate a relation between age and manner of shift. Kendler, Kendler, and Wells (1960) showed that nursery-school children, unlike adults, found an NRS easier than an RS. Kendler and Kendler (1959) found no difference at the kindergarten level in the relative ease of learning RS and NRS. Using a technique which allows *S* to respond in either a reversal or a nonreversal manner, Kendler, Kendler, and Learnard (1962) found that the percentage of children who responded in a reversal manner increased gradually from 37.5% at age 3 to 62.5% at age 10.

Thus, the data show that infrahuman organisms and children below the age of 5 or 6 execute an NRS more rapidly than an RS. The relative ease of the two types of shift is approximately equal for children 5–7 yr. old. After age 6 the relative ease of and/or preference for RS increases with age.

The present experiment seeks to assess the role in shift behavior of previous experience with the dimensions of the discrimination task, and it derives from a differentiation theory of discrimination learning. Present accounts of the comparative data on shift behavior deduce relative ease of RS from the assumption that mediating responses have been developed which serve to bring choice behavior in a discrimination task more firmly under the control of the stimulus dimensions (Zeaman & House, 1963) or of response-produced stimulation specific to the dimensions (Kendler & Kendler, 1962) rather than under the control of the specific stimulus values. Since the relevant dimension of a discrimination remains relevant in an RS but becomes irrelevant in an NRS such dimension-specific mediating responses should facilitate the learning of the former type of shift but interfere with the learning of the latter. An application of differentiation theory (Gibson, 1963; Gibson & Gibson, 1955) to this problem, however, suggests that increased sensitivity to stimulus variables themselves may play an important role in promoting the dimensional control of discriminative response which is assumed to be necessary for rapid reversal. According to this viewpoint, organisms may be expected to differ both individually and developmentally in their ability to detect stimulus dimensions and to independently utilize such invariant properties of stimulation to specify differences among objects. Such differences are assumed to stem, in turn, from differences in the organism's opportunities, learning ability, and perceptual experience. Thus, a condition which increases the likelihood that *S* will isolate and utilize the stimulus dimensions within a discrimination task may increase the ease of RS relative to NRS. This relation was tested in the present experiment by giving 5- and 6-yr.-old children nonreinforced pretraining designed to emphasize the independence and the dimensional nature of the properties of stimuli used in subsequent discrimination shifts.

---

[1] This paper is based on a dissertation submitted to Cornell University in partial fulfillment of the requirements for the PhD degree. The author is grateful to the members of her special committee, J. J. Gibson, A. L. Baldwin, and T. A. Ryan, and to T. J. Tighe for their helpful advice and criticism. She is also grateful to John Maguire, Superintendant of Schools, North Providence, Rhode Island, and to the principals and teachers in this school system for their generous cooperation.

[2] Now at Dartmouth College.

METHOD

Subjects

The *S*s were first-grade children from 67 to 84 mo. in age. Eighty-four *S*s were placed in the experimental (E) group and 104 in the control (C) group. Each *S* was run individually. Three *S*s, two in the E and one in the C group, failed to complete the experiment in one session and so were replaced.

Procedure and Apparatus

Pretraining stimuli and apparatus.—The stimuli were cylindrical aluminum tumblers which varied in height (Ht) and brightness (Br). They were hollow and open at one end. Brightness was varied by covering the tumblers with enamel paint. Each tumbler was 9 in. tall and $2\frac{1}{2}$ in. in diameter, but the apparatus used was so constructed that *E* could vary the Ht of the tumbler visible to *S*. The apparatus was constructed around a 32-in. high table with a 9 × 28 in. top. A 28×35 in. piece of plywood attached at a right angle to the front of the table shielded the table and motions of *E* from *S*'s view. A sliding door in the top center of the shield could be opened to reveal a three-walled compartment 8 in. deep with walls 8 in. wide × 16 in. high. The rear wall of the compartment was a sliding door which could be manipulated by *E* from behind the apparatus. A metal rod projected up through the center of a circular opening $2\frac{5}{8}$ in. in diameter in the center of the compartment floor. When the stimuli were inverted and placed on the rod, they extended through the hole and below the floor of the compartment. By manipulating a set screw beneath the floor of the compartment *E* could raise or lower the rod, thus changing the Ht of the tumbler visible to *S*. The sliding doors enabled discrete presentation of the stimuli.

E group pretraining.—In order to make clear to *S* the meaning of the instructions he was given in pretraining he was first given about 5 min. practice in making corrected same-different judgments on cookie cutters varying in shape and hue. Pretraining proper was then begun in which *S* was presented with a standard (St) tumbler of a given Ht and Br and was then required to render a same-different judgment on a series of successively and randomly presented comparison (Co) tumblers. None of these judgments were reinforced or corrected and all were based solely on vision. The instructions were as follows. An St was presented and *E* said, "Now let's pretend that this object is yours.

Look at it carefully. I am going to show you some other objects one at a time. You tell me which of the other objects that you will see are exactly like yours." The St was then removed, a Co stimulus presented and *E* said, "Is this object exactly like yours?" Then that stimulus was removed, a new Co stimulus presented, and the question was repeated. This procedure was followed for all of the Co stimuli. When an St was presented again to begin a new series the first sentence of the instructions was changed to, "Here is your object again." When a new St was introduced the first sentence of the instructions was, "Now let's pretend that this object and only this one is yours."

In the first stage of pretraining, which lasted about 5 min., *S* made judgments to stimuli which were easily discriminated as compared to those used in the second stage, and the Co stimuli varied only in Ht *or* in Br. Each *S* received only one of the following four Sts: a black (B) tumbler 4¾ in. high, a B tumbler 6¼in. high, a white (W) tumbler 4¾ in. high, a W tumbler 6¼ in. high. The Co stimuli were selected from tumblers of the following Ht and Br values: 3¼ in., 4¾ in., 6¼ in., 8¼ in., B, dark gray (DG—2 parts B to 1 part W), light gray (LG—1 part B to 50 parts W), W. Each *S* was given a series of Br variations and of Ht variations of the St with a maximum of four presentations of each type of series. Training in the first stage was terminated before this point if *S* made one errorless series of judgments of both Ht and Br differences.

The second stage of pretraining differed from the first in that the Co stimuli varied simultaneously on the two dimensions and the Co stimuli consisted of tumblers with the following values of Ht and Br: 4 in., 4¾ in., 6¼ in., 7 in., B, DG (1 part W to 6.5 parts B), LG (1 part B to 124 parts W), W. Each St was followed by a series of 16 Co stimuli which represented all possible combinations of these values. Each *S* made two series of judgments with each of the four Sts. The possible 24 orders of the four Sts were randomly distributed among *S*s within the restriction that for ½ the *S*s the last St presented would be identical to one of the positive stimuli in the later discrimination and for the other ½ the last St would be identical to one of the negative stimuli. In order to maintain *S*'s interest the St was re-presented in the middle of each series, and the number of stimuli which were identical to the St varied from one to three on any one series. The entire pretraining took about 30 min.

C group tasks.—The purpose of these tasks was to equate E and C *S*s in total exposure to *E* and other aspects of the experimental situation. The tasks were modifications of the Picture Completion and Picture Arrangement subtests of the Wechsler Intelligence Scale for Children. The modifications consisted of (*a*) presenting all 20 pictures in the completion test to each *S* and (*b*) requiring each *S* to tell a story about each picture in the arrangement test. None of *S*'s responses in either task were reinforced or corrected.

Discrimination task.—Discrimination training was begun for E and C *S*s immediately following the end of pretraining and the control tasks, respectively. All *S*s learned an initial discrimination and a shift of that discrimination. The specific procedure, stimuli, and apparatus were identical for both groups and replicated the conditions of Kendler and Kendler (1959).

The stimuli were four tumblers which were identical to the Sts used in pretraining, and the stimuli presented in the initial discrimination on any one trial consisted of one of the following two pairs: (*a*) the tall (T) white (W) tumbler vs. the short (S) black (B) tumbler, or (*b*) the TB tumbler vs. the SW tumbler. The *S* was instructed to pick up one of the tumblers on each trial and was told that if his choice were correct he would find a marble under it. The *S* was told that the "game" was to see how soon he could find a marble every time he chose. One half of the E and C *S*s were given an Ht discrimination with S as the positive stimulus and the other ½ were given a Br discrimination with B as the positive stimulus. When *S* made 9 correct out of 10 successive responses, the discrimination shift was presented immediately and continued to the same criterion with no change of instructions or interruption of procedure. Each of the E and C groups was divided equally into a reversal and a nonreversal group. That is, half of the E and C *S*s who had either S or B as the positive stimulus in the initial task now had T and W, respectively, as the positive stimulus (RS). The other half of the E and C *S*s who had either S or B as the positive stimulus in the initial task now had Br and Ht, respectively, as the relevant dimension (NRS). In the NRS equal numbers of *S*s had B, W, S, and T as the positive stimulus. The same tumblers used in the initial task continued to be used as stimuli during the shifts but were now presented with the irrelevant dimension held constant to minimize possible partial-reinforcement effects for *S*s undergoing NRSs.

Throughout the training the stimuli were presented according to a prearranged irregular sequence and each *S* was assigned to groups and positive stimuli on a predetermined random basis. At the end of the experiment *S* was allowed to choose a prize from an assortment which included charms, bubble gum, whistles, Hershey bars, peanuts, and raisins.

Word-association test.—As a check on whether pretraining altered the probability that the E group would employ relevant verbal labels, 50 E and 50 C *S*s were given a word-association test at the end of the discrimination shifts. The following words were presented in a random order: BLACK, WHITE, TALL, SHORT, NIGHT, TREE, SLEEP, BOY. Each *S* was asked to say the first word which came to mind when he heard the stimulus word.

RESULTS

An analysis of variance for repeated measurements showed that pretraining produced an equivalent effect on *S*s who were later placed in the reversal (ER) and in the nonreversal (ENR) groups ($F < 1$). Table 1 shows the errors in pretraining for these *S*s. In both groups there was a reduction in the total number of errors on the second of the two series of judgments which were made to each of the four successively presented Sts. The probability of these eight comparisons being in the same direction by chance is .004. Seventy-four percent of the errors in all Series I judgments were made to Co stimuli which were identical to the St in one dimension and differed by one or more steps on the second dimension. Seventy-two percent of the errors reduced from all Series Is to all Series IIs were also concerned with this type of response.

Although both trials and errors were recorded in both original learning and shift, only the former measure will be reported here. (Analyses of errors provided statements identical to those which result from analyses of trials.) The mean trials to criterion in original learning for the four major groups were: ER, 26.4; ENR, 27.3; CR, 27.4; CNR, 27.7. The mean trials to criterion for *S*s who had S as the positive stimulus and for those who had B as the positive stimulus were 27.3 and 27.5, respectively. A $2 \times 2 \times 2$ analysis of variance carried out on these data supports the conclusion that at the end of original learning the treatment groups were from the same population with respect to this discrimination problem and that the dimensions were of equal difficulty ($F < 1$).

Pretraining facilitated performance of reversal shift but did not affect performance of nonreversal shift. Table 2 shows the mean trials to criterion and *SD*s during shift for each group. A $2 \times 2 \times 2$ analysis of variance was used to evaluate the effects of pretraining, shift, and dimension. A $\sqrt{X + .5}$ transformation applied to the raw data reduced but did not eliminate heterogeneity of variance. Therefore, the 2.5%

TABLE 1

TOTAL NUMBER OF ERRORS ON EACH SERIES (S1 and S2) OF JUDGMENTS MADE TO THE SUCCESSIVELY PRESENTED STANDARDS IN PRETRAINING

Group	First Standard Judged		Second Standard Judged		Third Standard Judged		Fourth Standard Judged	
	S1	S2	S1	S2	S1	S2	S1	S2
ER	98	50	78	72	85	69	71	64
ENR	98	53	78	69	89	75	75	62

TABLE 2

MEANS AND *SD*s OF TRIALS TO CRITERION IN DISCRIMINATION SHIFTS

Group	*M*	*SD*
ER	5.6	6.77
ENR	17.1	19.46
CR	15.2	19.65
CNR	12.7	17.63

level was required for significance in the analysis of variance in order that the risk of a Type I error not exceed the 5% level. The results of the analysis, summarized in Table 3, showed that pretraining significantly interacted with type of shift. Dimensions also differed significantly but none of the other sources of variance reached significance. A separate analysis indicated that the performance of *S*s who had either Ht or Br as the relevant dimension was not influenced by which of the two values used for each dimension was the positive stimulus ($F < 1$ for each dimension).

The comparisons of specific group means were made by the Cochran-Cox test (Cochran & Cox, 1950, pp. 92–93). The procedure outlined by Ryan (1959) was used in assigning a level of significance to the comparisons of means. Since there were a total of six possible comparisons, a p level of .008 was required for each comparison to reduce the error rate per experiment to the .05 level. These comparisons, presented in Table 4, indicate that the ER group performed significantly faster than the ENR and CR groups. There was no difference between the CR and CNR groups.

TABLE 3

ANALYSIS OF VARIANCE OF TRANSFORMED TRIALS TO CRITERION IN DISCRIMINATION SHIFTS

Source	*df*	*MS*	*F*
Type of Shift (S)	1	7.77	2.00
Pretraining (P)	1	5.87	1.51
Dimension (D)	1	38.89	10.02***
S × P	1	33.53	8.64***
S × D	1	2.36	
D × P	1	14.03	3.61
S × P × D	1	.28	
Within Groups	180	3.88	

*** $p < .005$.

TABLE 4

COCHRAN-COX *t* STATISTICS AND OBTAINED PROBABILITIES FOR COMPARISONS OF DISCRIMINATION SHIFT GROUPS

Comparison	*t*	*p*
ER vs. ENR	3.25	$.002 < p < .005$
ER vs. CR	3.40	$.001 < p < .002$
ENR vs. CNR	1.06	$.20 < p < .30$
CR vs. CNR	0.85	$p > .30$

Note.—This is a two-tailed test.

In the word-association test the percentage of antonyms given as responses to the key words (i.e., BLACK, WHITE, TALL, SHORT) was 24.5 for *S*s in the E group and 25.4 for *S*s in the C group.

DISCUSSION

In discussing these results it should first be noted that the finding of no difference between the shift performances of the CR and CNR groups confirms Kendler and Kendler's (1959) data obtained with *S*s of a similar age range. Furthermore, the fact that the CR and CNR groups did not differ argues that the differential effects of the pretraining upon discrimination shifts is not due to differential effects of nonspecific transfer variables (e.g., warm-up). Consequently, any attempt to explain the selective facilitation of reversal shift performance in the perceptually pretrained group must deal with effects specific to the pretraining procedure.

Differentiation theory suggests that the observed facilitation resulted from an improvement during pretraining in *S*'s ability to detect and independently utilize the distinguishing features (dimensions) of the stimuli, i.e., from an increase in sensitivity to the stimulus variables themselves. The *S*'s performance during pretraining is consistent with this view. The nature of the pretraining errors indicates that the most difficult judgment was one which required *S* to use variations in a single dimension to specify differences, and that most of the learning which took place in pretraining was, in fact, an improvement in *S*'s ability to make such judgments. According to differentiation theory such improvement in sensitivity should increase the likelihood that *S*s would discriminate the objects of the discrimination task on the basis of the stimulus dimensions involved. In contrast, the control *S*s should be more apt to respond to the stimulus objects as undifferentiated wholes (e.g., the "large black" object vs. the "small white" object). Relative ease of RS may be deduced for an *S* who discriminates on the basis of the relevant dimension since it remains relevant in an RS and he has only to learn a relation between one aspect of this feature and reinforcement, while in an NRS he must first redirect his attention to (and perhaps further differentiate out) the formerly irrelevant dimension before learning the appropriate stimulus-reward relationship. On the other hand, relative ease of NRS may be deduced for an *S* who has reached criterion in original learning on the basis of perceived object-reward relationships, since only one of these relations is changed in such a shift, while both are changed in an RS.

Such an analysis, however, would not account for the fact that pretraining had no effect upon original learning. A hypothesis which may account for this finding is that the facilitating effect of having isolated the stimulus dimensions of a discrimination task will depend upon the number of dimensions involved. As the number of dimensions increases arithmetically the number of pairs of stimuli necessary to represent the variations will increase geometrically. Consequently, the value of being able to discriminate on the basis of dimensions rather than on the basis of the stimulus configurations themselves should become increasingly evident as the number of dimensions in the task increases. However, in the present experiment only two dimensions were involved and so only two pairs of stimulus objects were required to represent all variations. Thus, although pretrained *S*s are assumed to have been more likely than controls to have discriminated on the basis of dimensions during original learning, their problem would still be approximately equal in difficulty to that of the controls.

Explanations of the data on the RS-NRS comparison which emphasize the role of dimension-specific mediating responses might also claim the data of the present experiment if it is assumed that the appropriate mediating responses were developed under the pretraining conditions employed. In regard to the possibility of the establishment of verbal-mediating responses, the identical performance of control and experimental *S*s on the word-association test can be taken as an indication, although not as an unequivocal test, that pretraining did not alter the probability that experimental *S*s would employ relevant verbal labels during the discrimination tasks.

This study does not enable a conclusive statement as to the relative contributions of perceptual vs. mediational factors to the observed facilitation of reversal behavior. But the consistency of the operations and data of the present experiment with differentiation theory warrants the view that further consideration should be given to the possibility that differentiation processes play an important role in developmental changes in shift behavior.

REFERENCES

Brookshire, K. H., Warren, J. M., & Ball, G. G. Reversal and transfer learning following overtraining in rat and chicken. *J. comp. physiol. Psychol.*, 1961, **54**, 98–102.

Buss, A. H. Rigidity as a function of reversal and nonreversal shifts in the learning of successive discriminations. *J. exp. Psychol.*, 1953, **45**, 75–81.

Cochran, W. G., & Cox, G. M. *Experimental designs.* New York: Wiley, 1950.

Gibson, J. J. The useful dimensions of sensitivity. *Amer. Psychologist*, 1963, **18**, 1–15.

Gibson, J. J., & Gibson, E. J. Perceptual learning: Differentiation or enrichment? *Psychol. Rev.*, 1955, 62, 32–41.

Harrow, M., & Friedman, G. B. Comparing reversal and nonreversal shifts in concept formation with partial reinforcement controlled. *J. exp. Psychol.*, 1958, **55**, 592–597.

Kelleher, R. I. Discrimination learning as a function of reversal and nonreversal shifts. *J. exp. Psychol.*, 1956, **51**, 379–384.

Kendler, H. H., & Kendler, T. S. Vertical and horizontal processes in problem solving. *Psychol. Rev.*, 1962, 69, 1–16.

Kendler, T. S., & Kendler, H. H. Reversal and nonreversal shifts in kindergarten children. *J. exp. Psychol.*, 1959, **58**, 56–60.

Kendler, T. S., Kendler, H. H., & Learnard, B. Mediated responses to size and brightness as a function of age. *Amer. J. Psychol.*, 1962, **75**, 571–586.

Kendler, T. S., Kendler, H. H., & Wells, D. Reversal and nonreversal shifts in nursery school children. *J. comp. physiol. Psychol.*, 1960, **53**, 83–88.

Ryan, T. A. Multiple comparisons in psychological research. *Psychol. Bull.*, 1959, **56**, 26–47.

Tighe, T. J. Reversal and nonreversal shifts in monkeys. *J. comp. physiol. Psychol.*, 1964, **58**, 324–326.

Zeaman, D., & House, B. J. The role of attention in retardate discrimination learning. In N. R. Ellis (Ed.), *Handbook of mental deficiency.* New York: McGraw-Hill, 1963. Pp. 159–223.

(Received June 15, 1964)

Journal of Experimental Psychology
1965, 70, No. 4, 386–393

APPROXIMATIONS TO ENGLISH (AE) AND SHORT-TERM MEMORY: CONSTRUCTION OR STORAGE?[1]

ROY LACHMAN AND ABIGAIL V. TUTTLE
State University of New York at Buffalo *Hollins College*

Free recall improves with increasing contextual constraints on verbal material. 3 experiments sought to localize the facilitation at stimulus reception, storage, and/or retrieval. The construction hypothesis assumes associative and grammatical reconstruction of a training list during free recall: a retrieval effect. Experiment I compared retention for high and low AE lists on free recall and on a "successive binary recognition test" which precluded output construction. Result, high AE words are stored more efficiently. Experiment II tested for possibility of input construction during faulty stimulus reception as an alternative explanation to storage. Measurements of perceptual error during training failed to support the input-construction hypothesis. Evidence for output construction was slight but supported a prediction of instructional control of output-construction behavior. Experiment III demonstrated reproducibility and generality of results. Conclusions: (a) the human storage system is biased for high AE material, (b) recognition errors are lawful, false-positive error means are proportional to number of distracters, yielding a .19 constant.

In his 1942 book McGeoch proposed that there may be a high positive correlation, or even a perfect one, between meaning and the retention of verbal material. Meaning was undefined but psychologists generally agreed on the relative meaningfulness of specific verbal lists. Prose or poetry, e.g., was held to be more meaningful than a list of random words. In 1950 Miller and Selfridge questioned the prevailing interpretation. It was not meaningfulness per se that influenced retention, Miller argued, but rather meaningful material retained the statistical structure of the language. It was the latter, the order of approximation to English (AE) of a verbal list that was influencing retention. Miller and Selfridge (1950) demonstrated that the number of words recalled increases in a negatively accelerated fashion with order of AE. Although the precise form of the function has been questioned (Coleman, 1963), the pervasiveness of the general relationship is amply documented (Deese & Kaufman, 1957; Postman & Adams, 1960).

Deese (1961) recently suggested that superior retention of contextually constrained material may be due, in part, to "constructions" based on *S*'s prior learning of the associative and grammatical structure of the language. The construction hypothesis has been proposed in one form or another by a number of distinguished psychologists. Bartlett (1932) presented it in his classic work on memory. Woodworth (1938) stated, "In attempting to reproduce a story from memory one really *constructs* a story [p. 70]." More recently, Postman (1964) employed the hypothesis to integrate the results of a number of verbal-learning experiments.

---

[1] The research was supported by National Science Foundation Grant GB1587 and Undergraduate Research Participation Program G22791. Abigail Tuttle collaborated in Exp. I and II as a student participant of that program at Hollins College. Arnold Markowitz assisted in collection of data in Exp. III and IV.

Contamination of measures of forgetting by differences in degree of learning (storage) has been demonstrated by Underwood (1964). It should be added that the typical free-recall score may be a composite of not less than four factors: amount learned, amount forgotten, amount retrieved, and amount constructed. An alternative or adjunct hypothesis to construction is that a high AE list is more efficiently stored following input as well as or instead of construction during output. Order of AE may produce higher measures of retention at least in four ways: (*a*) it may produce more efficient storage of individual lexical units; (*b*) it may improve the accuracy of perception by permitting construction during input; (*c*) it may permit relevant construction during free recall; and (*d*) it may facilitate any combination of these factors.

EXPERIMENT I

The first experiment was designed to test the output-construction hypothesis by comparing the results of a test of free recall with recognition tests for material presented at two levels of AE. One group of *S*s received an English paragraph and a second group the same words in random order. Some *S*s received a free-recall test, the rest were given a recognition test. A two-choice recognition test for each word of a sequence of words, if properly paced, would seem to eliminate all opportunity for construction. Words of the initial list were randomly mixed with distracter words. This test list was presented one word at a time. The *S* simply sorted each word into a Yes or No category depending upon whether the word was or was not recognized as having been on the training list. There is neither the time nor the context necessary for construction under such circumstances. If superior retention for material of high AE *now* disappears, it would be safe to conclude that the locus of AE effects is essentially in the output system and that construction is a plausible explanation for superior retention. If the paragraph produces higher retention scores on recognition tests, it is reasonable to conclude that somehow the paragraph is getting into the brain more efficiently than random words.

Method

Design.—A 4 × 2 factorial design was employed involving two orders of approximation to English (high AE and low AE), and four conditions of retention testing (100, 200, or 400 distracters in recognition, and free recall).

Subjects.—The *S*s were 120 female undergraduate students at Hollins College ranging in age from 18 to 22 yr. Two additional *S*s were discarded.

Verbal material.—The learning material consisted of two 104-word lists containing 100 different words selected from the 1,000 most frequent words in the Thorndike and Lorge (1944) count or their derivatives (i.e., the inflectional suffix of some of the words was altered). Four function words occurred twice. Several paragraphs were constructed by students in accordance with the preceding specifications. The paragraph judged to conform most to grammatical convention was selected. The 104 words were presented one at a time. The word order was random for the low AE groups. The following order of presentation was employed for the high AE groups: "LATE LAST NIGHT ALTHOUGH LOSING MUCH SLEEP MARY AND I ARRIVED. . . ."[2]

Distracter words for the recognition groups were randomly selected likewise from the 1,000 most frequent words. The 100 different training words were randomly intermixed with either 100, 200, or 400 distracter words to produce recognition lists of 200, 300, and 500 words.

Procedure.—Simultaneous auditory and visual presentation of the words was employed for all groups during training and

---

[2] The training list is available from the author upon request.

Method

Design.—A 2 × 2 × 2 factorial design was employed involving high or low AE (paragraph or random word order), overt or covert articulation of the stimulus words; and instructions to construct or neutral instructions.

Subjects.—The *S*s were 64 undergraduates of both sexes attending Roanoke College summer session and ranging in age from 16 to 24 yr. Three *S*s were discarded and replaced.

Procedure.—Training words were identical to those in Exp. I but were now presented on a DuKane filmstrip viewer. Training was restricted to the visual modality and words appeared one at a time at 1.5-sec. intervals. The *S*s were run individually and the testing was exclusively free recall. The construction groups were told to attempt to fill up the test sheet even if they had little confidence in some of the words that they would be writing. They were also instructed to attempt to figure out what words had been previously presented, to check the words that they reported with confidence, and to put an "X" next to the low-confidence words. The "nonconstruction" groups were presented instructions similar to those of Exp. I. The overt articulation groups were directed to read the training words out loud. This instruction was omitted for the covert groups.

Results

Inflectional suffixes were again ignored in counting the number of words correct. AE is highly significant, F (1, 56) = 46.69, $p < .001$, with mean words recalled for high and low AE, respectively, 27.66 and 17.06. Instructions are significant, F (1, 56) = 4.83, $p < .05$: the mean is 24.06 for instructions to construct, and 20.66 for neutral instructions. There was no evidence of perceptual error for the overt articulation *S*s; the mean words correct are virtually identical for the overt and covert groups, 22.72 and 22.00, F (1, 56) = .22. All interactions, except AE ×Instructions, yielded $p > .10$. The AE × Instructions interaction is marginal, F (1, 56) = 3.51, $.05 < p < .10$. Orthogonal comparisons, however, confirm the prediction as directions to construct facilitated recall only for the high AE group. Mean words correct is 30.81 for construction instructions, and 24.00 for neutral instructions when the learning material is high AE, t (56) = 2.93, $p < .01$. The means for the low AE list are 17.61 for instructions to construct, and 16.81 for neutral instructions, t (56) = .35.

It will be recalled that the construction *S*s were told to denote their high- and low-confidence words. Mean words correct and incorrect in accordance with this distinction are presented in Table 3. The high- and low-confidence words were analyzed independently. AE is significant both for high-confidence words, F (1, 28) = 20.74, $p < .001$, and for low-confidence words, F (1, 28) = 5.67, $p < .05$. Both sets of treatment effects for verbalizations and both sets of interactions yielded $p > .10$.

TABLE 3

MEAN WORDS CORRECT AND INCORRECT FOR OUTPUT-CONSTRUCTION GROUPS: EXP. II

Approximations to English	High Confidence Words Correct	Low Confidence Words Correct	High Confidence Words Incorrect	Low Confidence Words Incorrect
High	27.44	3.38	3.19	4.75
Low	16.00	1.31	2.81	3.44
Verbalization:				
Overt	21.38	2.88	3.13	4.75
Covert	22.06	1.81	2.89	3.44

EXPERIMENT III

Although the results of the previous experiments are unambiguous, generality has not been established since a single sample of verbal material was employed. Experiment III is a partial replication of the first experiment employing the initial verbal material and three additional lists. The words were presented in high or low AE and the tests administered were recognition with 100 distracters or free recall. Eight different samples of distracters were used.

Method

Design.—A 4 × 2 × 2 factorial design was employed involving high or low AE, free recall or a recognition test, and four sets of verbal materials.

Subjects.—The *S*s were 240 introductory psychology students (male and female) at Buffalo. Five *S*s were discarded and replaced.

Verbal material.—Four sets of training words were employed.[3] A different random order of words was presented to each *S* in the low AE condition. Eight different samples of 100 distracter words were selected from approximately 600 words not used in the training lists. All distracter words, thus, were from the 1,000 most frequent words in the Thorndike-Lorge (1944) count. The order of training and distracter words was random and was different for each *S* in the recognition condition. Each word was printed by an IBM 557 on the left center of a blank IBM card, one word to a card.

Procedure.—The *S*s were run in groups of 16. Each *S* was randomly assigned to 1 of the 16 treatments. Two additional blocks were run to replace discarded *S*s or those that failed to show up. The same general instructions were administered but each *S* was told the character of his training list, either a paragraph or random words. Pretraining consisted of 30 sec. of practice turning of blank IBM cards. Training was restricted to the visual modality; *S* turned his IBM cards and exposed each word to a 1.5-sec. beat played from magnetic tape. Roughly 10% of the *S*s did not keep pace with the beat. The deviation, for these *S*s, from the assigned total exposure time was never more than 25 sec. The interval between training and testing ranged between 5 and 45 sec. for individual *S*s. The recognition *S*s sorted the cards to a 2.5-sec. beat. The slowest *S* required 50 sec. more than the fastest. In all other respects, the procedure was the same as in Exp. I.

Results

The adjustment of recognition scores and the scoring of free recall are identical to that of Exp. I. Cell means for each of the 16 treatments are presented in Table 4. Superior retention of high AE material occurred for each sample of verbal material in free recall and recognition. The results of Exp. I, thus, were produced for the list of Exp. I and for three new samples of verbal materials. The analysis of variance for these data produced the following results, retention tests: $F\ (1, 224) = 282.94,\ p < .001$; AE: $F\ (1, 224) = 41.60,\ p < .001$; lists: $F\ (3, 224) = 4.12,\ p < .01$. All interactions yielded $p > .10$.

The overall mean for false-positive errors is 19.3 for high AE and 19.0 for low AE. Mean false-positive errors for the high AE lists are 16.7 for List A, 16.9 for List B, 19.9 for List E, 22.5 for List F, and for the low AE lists 20.5 for List A, 16.7 for List B, 22.2 for List E, and 17.7 for List F. Inasmuch as 100 distracters were

---

[3] A fourth experiment ($N = 120$) was conducted. Five additional paragraphs were selected from lists constructed in accordance with the criteria presented in Exp. I. The six lists were administered in high AE to six groups of *S*s, followed by a free-recall test. Mean number of words recalled for each paragraph is,

(A) 22.30 (B) 25.10 (C) 26.85 (D) 30.55 (E) 31.25 (F) 35.70

Means *without* common lines are significantly different (Duncan range test, $\alpha = .05$). Lists A, B, E, and F were administered in Exp. III; List E was previously employed in Exp. I and II. The lists are available from the author upon request.

TABLE 4

TREATMENT MEANS: EXP. III

List	Topic	Free-Recall Test		Recognition Test		Collapsed
		High AE	Low AE	High AE	Low AE	
A	SHIP	22.07	19.53	49.13	37.27	32.00
B	GARDEN	26.73	20.07	47.07	44.07	34.48
E	TRIP	27.53	19.53	43.73	35.67	31.62
F	WAR	31.47	24.53	51.73	38.93	36.67
Collapsed:		26.95	20.92	47.92	38.98	

utilized, the ratio of false-positive errors is .19 for both high and low AE and is remarkably uniform.

DISCUSSION

The effect of an experimental arrangement upon retention typically is measured by the quantity of stimulus material reproduced. Results so obtained may grossly reflect the influence of an independent variable on one or more discrete processes such as the quantity and character of stimulus *reception*, *storage*, or *retrieval*. A verbal stimulus signal transmitted to an *S* may be received accurately, inaccurately, or not at all. A stimulus received may be stored permanently, temporarily, or not at all. A stored signal may be retrieved efficiently, inefficiently, or not at all. Contextual constraint clearly increases the quantity of verbal material subsequently reproduced. It has not been clear, however, whether high AE facilitates reception, storage, or retrieval.

Increased accuracy of stimulus reception is possible if a high AE list generates input construction; Exp. II demonstrated no such effect. Immediate reproduction was obtained by overt verbalization during stimulus input. Both high and low AE verbal material produced completely veridical stimulus reception: No articulation errors were observed. Nor did any difference in retention result from covert and overt verbalization. Input construction is simply not necessary for verbatim reception of the high-frequency words presented.

Experiments I and III reliably demonstrated that high AE material is more efficiently stored. Three quantitatively different recognition tests were administered in Exp. I. Four samples of verbal material were presented in Exp. III. In each of the seven treatments substantially more high AE words were recognized. The recognition tests precluded output construction and Exp. II ruled out input construction. There are few reasonable alternatives to the conclusion that the human storage system is biased for contextually constrained material. The logic involved in extending conclusions based on recognition behavior to free recall requires comment. Recognition and free recall are essentially two different retrieval processes. A variable that facilitates storage will do so regardless of the subsequent retrieval system employed. The several tests of recognition demonstrate that high AE material is stored more efficiently. If this is granted, it follows that superior free recall of a high AE list reflects superior storage and presents no problem. Indeed, if a high AE list failed to enhance free recall, a problem would now exist to account for such an anomaly.

The preceding reasoning in no way precludes output construction as an additional factor. Separate instructions were administered in Exp. II in an attempt to generate construction. The result was a higher level of free recall selectively for high AE material. A difficulty remains as the instructions may have merely produced more selection from a larger storage pool. Interrogation of *S*s and their behavior suggest that free recall involves at least two distinct processes. Short bursts of sequentially related words are emitted by *S* (Postman & Adams, 1960) in Phase 1 from what may be a short-term storage pool. Phase 2 may consist of a covert recognition process in which *S* implicitly selects successive words from an associative pool (Deese, 1962) primed by initial training, rejecting some words and emitting others.

Recognition measures of retention typically involve simultaneous presentation of a correct item with one or more distracters. One item is selected by *S* from each set (Murdock, 1963; Postman, Jenkins, & Postman, 1948). Specific techniques of distracter selection and matching with correct items frequently leads to serious methodological difficulties. The procedure developed for the present research could be called *successive binary recognition*. Distracters were randomly sampled from the same population as the correct words. Both types of items were randomly intermixed, presented one at a time, and judged as correct or incorrect. One result is strikingly lawful. The false-positive error mean is proportional to the number of distracters; the constant is very close to .19 if errors are averaged over a large sample. This constant in conjunction with the quantity of obtained errors in free recall will permit calculation of the number of words that *S* implicitly generates during free recall. If the procedure can be validated by a single independent estimate, a powerful tool will be available for determining implicit behavior.

REFERENCES

BARTLETT, F. C. *Remembering: A study in experimental and social psychology.* Cambridge: Cambridge Univer. Press, 1932.

COLEMAN, E. B. Approximations to English: Some comments on the method. *Amer. J. Psychol.*, 1963, **76**, 239–247.

DEESE, J. From the isolated verbal unit to connected discourse. In C. N. Cofer (Ed.), *Verbal learning and verbal behavior.* New York: McGraw-Hill, 1961. Pp. 11–31.

DEESE, J. On the structure of associative meaning. *Psychol. Rev.*, 1962, **69**, 161–175.

DEESE, J., & KAUFMAN, R. A. Serial effects in recall of unorganized and sequentially organized verbal material. *J. exp. Psychol.*, 1957, **54**, 180–187.

MCGEOCH, J. A. *The psychology of human learning.* New York: Longmans, Green, 1942.

MILLER, G. A., & SELFRIDGE, J. A. Verbal context and the recall of meaningful material. *Amer. J. Psychol.*, 1950, **63**, 176–185.

MURDOCK, B. B., JR. An analysis of the recognition process. In C. N. Cofer & B. S. MUSGRAVE (Eds.), *Verbal behavior and learning.* New York: McGraw-Hill, 1963. Pp. 10–22.

POSTMAN, L. Short-term memory and incidental learning. In A. W. Melton (Ed.), *Categories of human learning.* New York: Academic Press, 1964.

POSTMAN, L., & ADAMS, P. A. Studies in incidental learning: VIII. The effects of contextual determination. *J. exp. Psychol.*, 1960, **59**, 153–164.

POSTMAN, L., JENKINS, W. O., & POSTMAN, D. L. An experimental comparison of active recall and recognition. *Amer. J. Psychol.*, 1948, **61**, 511–519.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, 1944.

UNDERWOOD, B. J. Degree of learning and the measure of forgetting. *J. verbal Learn. verbal Behav.*, 1964, **3**, 112–129.

WOODWORTH, R. S. *Experimental psychology.* New York: Holt, 1938.

(Received June 16, 1964)

there were 10 practice sets, 1 each of the 10 set types. Each adjective that was used in the experimental sets for a given *S* appeared exactly once in the practice sets for that *S*.

Procedure.—Each *S* received two mimeographed booklets, one of practice sets, and one of experimental sets. Except for the cover sheet, each page of the booklet contained two adjectives on each of one or two lines, together with a blank space in which *S* recorded his response. For $\mathrm{HHM^+M^+}$ sets and $\mathrm{LLM^-M^-}$ sets, the adjectives appeared in alternating order which was balanced over the two sets of each type. The pages of each booklet were arranged in a different scrambled order for each *S*.

The *S* was instructed to read the adjectives slowly to himself. A total of 15 sec. was used for each set. At the end of the allotted time, *E* said, "Rate & Turn." The *S* then recorded his response and turned to the next page.

The *S*s were told to use 50 to rate a person they would neither like nor dislike, to use lower numbers for persons they would dislike, and to use higher numbers for persons they would like. Except for centering the scale at 50, *S*s were told they could use any numbers they wished. It was emphasized that they should feel free to change their use of the rating scale at any time during the practice sets.

Subjects.—The *S*s were 48 males who were fulfilling a course requirement in introductory psychology. They were run one or two at a time, and were randomly assigned in equal numbers to the four stimulus replications.

RESULTS

Table 1 lists mean response per set for each of the 10 set types. There are three main aspects of the data that are of interest.

TABLE 1

MEAN RESPONSE PER SET AS A FUNCTION OF SET TYPE

Set Type	Response	Set Type	Response
HH	72.85	HHHH	79.39
$\mathrm{M^+M^+}$	57.56	$\mathrm{M^+M^+M^+M^+}$	63.20
$\mathrm{M^-M^-}$	42.18	$\mathrm{M^-M^-M^-M^-}$	39.50
LL	23.70	LLLL	17.64
		$\mathrm{HHM^+M^+}$	71.11
		$\mathrm{LLM^-M^-}$	25.67

Note.—H, $\mathrm{M^+}$, $\mathrm{M^-}$, and L stand for Highly Favorable, Moderately Favorable, Moderately Unfavorable, and Highly Unfavorable adjectives. Response scale is centered at 50; higher numbers indicate favorable impressions, lower numbers indicate unfavorable impressions.

The primary purpose of the experiment was to get a comparative test of the additive and averaging formulations in the way alluded to in the first paragraph of the introduction. The critical comparison is based on the HH and the $\mathrm{HHM^+M^+}$ sets. Both sets have the same highly favorable adjectives, HH; the latter set has also the added moderately favorable adjectives, $\mathrm{M^+M^+}$. The addition of this moderately favorable information should increase the favorableness of the response if the stimuli are combined by an additive process. Thus, an additive model would predict that the response to $\mathrm{HHM^+M^+}$ would be higher than the response to HH alone. The observed difference is in the opposite direction: The mean response of 71.11 for $\mathrm{HHM^+M^+}$ is less than the mean response of 72.85 for HH.

The LL and $\mathrm{LLM^-M^-}$ sets give a similar comparison. Both sets have the same highly unfavorable LL adjectives, but the latter set has also the added moderately unfavorable $\mathrm{M^-M^-}$ adjectives. An additive model would thus predict a lower response to $\mathrm{LLM^-M^-}$ than to LL. The observed difference is again in the direction opposite to prediction: The mean response of 25.67 for $\mathrm{LLM^-M^-}$ is higher than the mean response of 23.70 for LL.

To test these results, each *S* was given a difference score for each comparison, the direction of difference being taken so that an additive model would predict a positive score. The overall mean difference of -1.86 was significantly less than zero, $F(1, 44) = 6.50$. The stimulus-replication factor did not approach significance, indicating that the results did no arise from a fortuitous choice adjectives. It is thus concluded that, for the present stimuli at least, the addition of moderately polarized information to highly polarized information decreases the polarity of the response. This conclusion, which contradicts a simple additive model, is consistent with an averaging formulation.

Since the critical result is qualitative, resting on a direction of difference, it cannot reasonably be attributed to a shortcoming in the scale of measurement. Moreover, adding moderately polarized to moderately polarized adjectives actually increases the polarity of the response. This can be seen in the comparisons between sets of two and four M^+ or M^- adjectives (Table 1, second and third lines). This set-size effect supports the interpretation of the critical result in two respects. First, it validates the nominal neutral point of the rating scale and the choices of moderately polarized adjectives. Second, it underscores the contrary direction of the effect in the critical comparison.

The second aspect of the results is the effect of set size. These data bear on the adequacy of an averaging formulation. In agreement with previous findings cited above, Table 1 shows that for adjectives of equivalent value, four adjectives produced a more polarized response than two adjectives. The mean magnitude of the difference, averaged over the four scale ranges, is 5.23; the corresponding mean standard error is 1.02.

This result would seem contrary to an averaging formulation. If the response equaled the mean scale value of the stimuli, it would be the same for sets of two and sets of four adjectives. However, the data may be accounted for by making use of an initial impression, analogous to the initial opinion employed in previous work with an averaging-type model for serial presentation (Anderson, 1959). It is assumed that S has an initial impression, I_0, which is combined in a weighted average with the scale values of the stimuli of each set. The response to a set of Size 1 is then given by

$$R_1 = wA + (1 - w)I_0, \qquad [1]$$

where A is the scale value of the single item, w is the weight associated with the item, and $1 - w$ the weight for the initial impression.

For sets of Size k in which all stimuli have the same weight and scale value, the equation for the response is

$$R_k = \frac{kwA + (1 - w)I_0}{kw + (1 - w)}. \qquad [2]$$

In this expression, the numerator is the weighted sum of the scale values, and the denominator is just the sum of the weights.

To apply these equations, the HH and LL sets were considered as of Size 1, with scale values of 100 and 0. I_0 was set equal to 50, the nominal neutral point of the response scale. Equation 1, together with each S's observed mean response to the HH or LL sets, was then used to estimate w for that S. Equation 2, with $k = 2$, was then applied to predict S's response to the sets of four adjectives. The mean predicted response to HHHH was 79.7 which is .3 higher than the obtained value; the mean predicted response to LLLL was 16.6 which is 1.0 lower than the obtained value. (These predictions differ from those obtainable from the group data of Table 1 because the mean ratio does not equal the ratio of the means.)

Although the agreement between predicted and obtained is fairly good, the fit should perhaps be taken as no more than a qualitative indication that the averaging formulation can account for the effect of set size. Unfortunately, it did not seem feasible to estimate the values of A for each S separately. The use of 0 and 100 as scale values for L and H adjectives, although seemingly appropriate for most Ss, did not appear to be valid for all. A few responses were small negative numbers and, more seriously, some Ss did not appear to be using a full 0–100 range for their responses. Although a center-anchored response scale was desirable for the main purpose of this experiment, an end-anchored scale will probably be preferable for a rigorous quantitative test of the averaging model explanation of the effect of set size.

The above analysis rests on the use of the initial impression, a concept that seems natural enough in the context of opinion formation where it was previously employed. Here, however, it might more reasonably be conceptualized as a neutral impression, based on a lack of information or on an expectation developed by previous sets, which is progressively changed as the stimuli of the present set are integrated into the impression.

The two aspects of the results considered so far concern essentially qualitative tests of the additive and averaging formulations. The third aspect of the data involves two quantitative comparisons for which both formulations make the same prediction: The response to HHM+M+ should equal the average of the responses to HHHH and M+M+M+M+; similarly, the response to LLM−M− should equal the average of the responses to LLLL and M−M−M−M−.

These predictions are straightforward. Suppose that all H adjectives have the same scale value, s_{H}, and that all M+ adjectives have the same scale value, s_{M^+}. Let $R(X)$ denote the response to a set, X, of stimuli. In a simple additive model, the response is just the sum of the scale values:

$$R(\mathrm{HHHH}) = 4s_{\mathrm{H}},$$
$$R(\mathrm{M^+M^+M^+M^+}) = 4s_{\mathrm{M}^+},$$
$$R(\mathrm{HHM^+M^+}) = 2s_{\mathrm{H}} + 2s_{\mathrm{M}^+}.$$

The last expression is seen to be the average of the first two, as asserted.

The averaging model makes the same prediction if all stimuli have the same weight. This may be shown by direct application of Equation 2. However, since all sets in question have equal size, the initial impression may be ignored and the above listed expressions simply divided by 4 to yield the same conclusion.

The above derivations are based on the simplifying assumption that all stimuli in a given scale range have the same scale value. Because of the balancing in the set construction, it is straightforward to show that the same predictions are obtained if unequal scale values are allowed.

The test of the prediction is made directly from the raw data. Each S was given a score obtained by summing his responses to the two HHHH sets and the two M+M+M+M+ sets, and subtracting his responses to the four HHM+M+ sets. According to prediction, this dependent variable has a true mean of zero. From Table 1, the observed mean is .19 on a per set basis, and this is not significantly different from zero. However, the observed mean for the corresponding comparison involving the L and M− adjectives was 2.90 which is significantly different from zero, $F(1, 44) = 8.53$.

This significant discrepancy naturally raises some doubt that either formulation can give an exact, quantitative account of the data. It may be worth noting, however, that the discrepancy is in the direction that would be predicted by Equation 2 if the M^- adjectives had lower weights than the L adjectives.

DISCUSSION

The main result is that the addition of moderately polarized stimuli to highly polarized stimuli decreases the polarity of the response. Any simple application of an additive formulation would predict that response polarity would increase as the number of stimulus items of like polarity is increased. The main result is clearly inconsistent with such a formulation.

An averaging formulation is, of course, consistent with the main result. The addition of moderately polarized stimuli to highly polarized stimuli decreases the mean polarity of the set. If *S* responds in terms of the mean scale value, his response should also then decrease.

But it should be noted that the averaging model does not imply that this result will necessarily be obtained. In the model, the response is considered as an average, not of the stimulus values alone, but includes also the initial impression. This average may itself increase even though the mean stimulus value decreases. It is, in fact, in just this way that the averaging model accounts for the effect of set size.

Although the present results appear to favor an averaging over an additive formulation, there are certain qualifications that need mention. First and most important, both models failed one of the two quantitative tests. The possibility of accounting for this discrepancy within the averaging model was noted above. However, the discrepancy might also stem from some methodological shortcoming, such as the lack of a theoretically adequate response scale (Anderson, 1962b). And the discrepancy may, of course, reflect some basic theoretical flaw in the models, such as changes in the meanings or scale values of the adjectives as a function of context. The present experiment does not, unfortunately, distinguish among these possibilities.

The second qualification is that the results may be peculiar to the adjective stimuli. In particular, if *S* judged the value of a bundle of commodities, as in Gulliksen (1956), the addition of even a slightly valued commodity would no doubt increase the value of the bundle. In that case, of course, an averaging model would not apply. This example suggests that perhaps an averaging formulation will be applicable to judgments of single more or less definite objects, in contrast to judgments of collections or bundles of objects. Added information would simply reveal more properties of the object in the former case, but would correspond to an actual change in the collection in the latter case.

REFERENCES

ABELSON, R. P. Do predispositional factors summate? *Amer. Psychologist*, 1961, **16**, 377. (Abstract)

ANDERSON, N. H. Test of a model for opinion change. *J. abnorm. soc. Psychol.*, 1959, **59**, 371–381.

ANDERSON, N. H. Application of an additive model to impression formation. *Science*, 1962, **138**, 817–818. (a)

ANDERSON, N. H. On the quantification of Miller's conflict theory. *Psychol. Rev.*, 1962, **69**, 400–414. (b)

ANDERSON, N. H. Primacy effects in personality impression formation using a generalized order effect paradigm. *J. pers. soc. Psychol.*, 1965, **2**, 1–9.

GULLIKSEN, H. Measurement of subjective values. *Psychometrika*, 1956, **21**, 229–244.

HAMMOND, K. R. Probabilistic functioning and the clinical method. *Psychol. Rev.*, 1955, **62**, 255–262.

JOHNSON, D. M. *The psychology of thought and judgment.* New York: Harper, 1955.

LEVY, L. H., & RICHTER, M. L. Impressions of groups as a function of the stimulus values of their individual members. *J. abnorm. soc. Psychol.*, 1963, **67**, 349–354.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning.* Urbana: Univer. Illinois Press, 1957.

PODELL, J. E. The impression as a quantitative concept. *Amer. Psychologist*, 1962, **17**, 308. (Abstract)

RIMOLDI, H. J. A. Prediction of scale values for combined stimuli. *Brit. J. statist. Psychol.*, 1956, **9**, 29–40.

SPENCE, W., & GUILFORD, J. P. The affective values of combinations of odors. *Amer. J. Psychol.*, 1933, **45**, 495–501.

STEWART, R. H. Effect of continuous responding on the order effect in personality impression formation. *J. pers. soc. Psychol.*, 1965, **1**, 161–165.

TRIANDIS, H. C., & FISHBEIN, M. Cognitive interaction in person perception. *J. abnorm. soc. Psychol.*, 1963, **67**, 446–453.

WEISS, W. Scale judgments of triplets of opinion statements. *J. abnorm. soc. Psychol.*, 1963, **66**, 471–479.

WILLIS, R. H. Stimulus pooling and social perception. *J. abnorm. soc. Psychol.*, 1960, **60**, 365–373.

(Received June 18, 1964)

Journal of Experimental Psychology
1965, 70, No. 4, 401–406

SPATIAL AND TEMPORAL DETERMINANTS OF VISUAL BACKWARD MASKING[1]

ROBERT W. SEKULER[2]

Brown University

Duration thresholds were obtained for a single test stripe succeeded by a masking pattern of alternating dark and bright bars delivered to the same retinal region. The effect of the angle (alpha) between test stripe and masking contours was studied in conjunction with both vertical and horizontal test stripes. With either of the test-stripe orientations duration threshold decreased as alpha increased from 0° to 90°. In another experiment both alpha and the interstimulus interval (ISI) were varied. The effectiveness of alpha as a determinant of masking seemed to be restricted to ISI below 60 msec.

When two brief flashes are presented in rapid succession to the same retinal area the detectability of the first is reduced. This phenomenon has been termed visual backward masking. The effect of the spatial relationship between the first (or "test") stimulus and the second (or "masking") stimulus has yet to receive adequate systematic analysis though several lines of evidence testify to its importance.

Battersby and Wagman (1962, 1964) have shown the important role played by the spatial relationship between the perimeters of otherwise contourless test and masking flashes. Kolers (1962; Kolers & Rosner, 1960) has extended the analysis using a concentric disk and ring as test and masking stimuli.

In order to understand the contribution of contours *within* the masking field one must turn to studies using spatially overlapping, patterned test and masking stimuli. In many of these studies, however, the amount of overlap between the successively presented contours cannot be precisely stated because of the stimulus configurations employed (Kinsbourne & Warrington, 1962a, 1962b; Schiller, 1965; Schiller & Wiener, 1963). All of them used letters of the English alphabet as test stimuli and the diversity of their form caused the test figures to be overlapped in a different fashion by the microstructure of the masking field. In addition, the masking patterns used do not facilitate the analysis of the specific contour relationships between stimuli.

A more precise statement of the role played by contour relationships in backward masking might be derived from experiments using geometrically simple test and masking contours. In the three experiments to be reported backward masking is measured by the duration threshold for a small dark stripe masked by a pattern of alternating dark and bright stripes. The particular spatial relationship selected for study is the angle (alpha) between the test stripe and the orientation of the masking contours. In Exp. I and II backward masking is measured for different values of alpha for both vertical and horizontal test stripes. In Exp. III both alpha and ISI are varied in an effort to study the interaction between these spatial and temporal determinants of backward masking.

---

[1] This paper is based on a portion of a dissertation submitted in partial fulfillment of the requirements for the PhD degree. The research was supported by a grant from the National Science Foundation (G-14314) to Harold Schlosberg and by a Public Health Service Predoctoral Fellowship. The author is indebted to the late Harold Schlosberg for his encouragement and advice and to Trygg Engen and Lorrin A. Riggs for their invaluable suggestions.

[2] Now at Northwestern University, Evanston, Illinois.

METHOD

Subjects.—The *O*s in Exp. I and II were five undergraduates. Following three practice sessions, they were paid at an hourly rate of $1.25. The four *O*s in Exp. III, none of whom had served in the two preceding experiments, were also undergraduates. These *O*s were paid the hourly rate plus a bonus of 1/10 cent for each correct response. All *O*s had corrected acuity of 20/20 or better.

Apparatus.—All observations were made in a darkened, sound-resistant chamber. The *O* viewed the targets from a distance of 200 cm., his head steadied by a fixed chin and headrest. Binocular viewing with natural pupils was used throughout. Centered about the middle of the target area was a white adapting screen subtending 16° of arc on the horizontal by 19° of arc on the vertical. In the center of the adapting background was a circular aperture with diameter subtense of 2°44′ of arc. Directly behind this aperture was a rear projection screen. Stimuli for the experiments were projected onto the back of this screen by an electronic projection tachistoscope.

The three-channel tachistoscope was located behind the rear projection screen with each channel consisting of a glow modulator tube (Sylvania, R1166) and optics suitable for focusing the interposed stimulus transparency upon the rear projection screen. Measurements with a calibrated photomultiplier tube indicated that the maximum difference between channels constituted 1.5–2.0% of total light output. Over the course of repeated observations flashes generated by the glow modulator tubes rose to 90% of maximum within 1 msec. Decay times were somewhat less.

The stimulus patterns to be presented were photographically reduced and imaged on glass photographic plates. Carefully machined channels in front of the glow modulators permitted a slide to be inserted, withdrawn, and reinserted with a negligible change in position.

The adapting field and circular target area were illuminated by a 300-w. projector located behind and above *O*. The luminances of the adapting screen and target area, measured with a Salford Electrical Instruments, Ltd. (SEI) photometer, were 10.12 mL. and 2.35 mL.

On each trial the initial flash contained either a small dark test stripe or a homogeneous flash of equivalent luminance. The test stripe subtended 57′ × 6′ of arc and its longer axis was vertical. The masking pattern, photographically reduced from a Ronchi ruling (Edmund Scientific Company, Barrington, N. J.) had bright and dark stripes of equal width, 6′ of arc. The contrast of the test stripe and that of the masking stripes was measured under full experimental conditions. These values were .14 and .28. The luminance of the circular area surrounding the test stripe was .65 mL. while the luminance of the center of each bright stripe in the masking pattern was 1.67 mL. A dot of 1.5′ of arc subtense was located in the center of the target area, providing a fixation point. The test stripe was centered about this point.

EXPERIMENT I

Procedure

Each session began with a minimum of 6-min. adaptation to the background screen at the luminance used in the experiment. Each trial consisted of a first (test) flash of variable duration, an interstimulus interval (ISI) of 40 msec., and a masking flash of 100-msec. duration. During ISI the photometric condition of the target area was the same as between trials.

Alpha, the angle between test stripe and masking contour orientation, was the independent variable. The following alphas were used: 0, 7.5, 15, 22.5, 30, 45, 60, 67.5, 75, 82.5, and 90°. Two different conditions prevailed with alpha = 0°. In one condition the region of the target area upon which the dark test stripe appeared was subsequently occupied by a section of dark stripe in the masking pattern; in the other this same position was occupied by a bright portion of the masking pattern.

Stimulus presentations were initiated by *O* at the termination of a .5-sec. auditory signal. The *O* then used a switch to indicate whether he had seen the test stripe on that trial. The interval between trials was approximately 15 sec., during which time *O* confined his gaze to the target area.

Each alpha to be studied was presented in a block of 25 trials. The test stripe was omitted on a preselected 7 of the 25; on these "catch" trials a homogeneous field of equivalent luminance was presented in place of the test-stripe field. Within the 25-trial block an "up-down" or tracking procedure was used to measure the duration threshold. On the first trial of any block the test stripe was presented at a duration of 7 msec. If *O* responded "No" (that he had not seen the stripe), the stripe was presented with increased duration on the next trial; if *O* responded "Yes" (that he had seen the stripe), the next presentation would be shorter. The *O*'s responses themselves controlled the signal duration, keeping it in oscillation about some central or threshold value.

Six different alphas were used in any session and over the course of the experiment an *O* was run in four sessions with each condition. The order of testing with the different alphas was individually randomized for each *O* under the restriction that no condition could be run $n + 1$ times until all other alphas had been used n times. Knowledge of results was not provided. The *O* had been informed that the test stripe would be omitted on some trials but was not told the frequency of its omission.

Results

The duration threshold was defined as the mean of the test-stripe durations associated with transition responses, shifts from Yes to No or vice versa. The results of Exp. I are summarized in Fig. 1. Each point plotted is the mean of the four threshold determinations for a single *O* and alpha value. The threshold for the detection of the black vertical test stripe decreases with increasing alpha. It will be recalled that two different conditions were studied at alpha 0° (see above). Since analysis showed the difference between thresholds for the two 0° alphas to be statistically nonsignificant—t (44) = 0.097— for simplicity of representation, the points plotted against alpha = 0° are mean values taken across both conditions. An analysis of variance indicated that both the alpha and Alpha × *O* effects were significant beyond the .01 level, F (11, 180) = 46.30; F (44, 180) = 1.77. In every case the normal or unmasked threshold was found to lie between 2.5 and 3.0 msec. below the threshold for the largest alpha values studied. This suggests that some small amount of backward masking is occurring even with the large alpha conditions. Tests of the significance of the difference between means of successive alphas indicated that beyond 45° the effects of increasing alpha were not statistically significant ($p > .05$; t test based on the alpha term with $df = 11$). This confirms the impression given by Fig. 1, that the function relating backward masking to alpha is composed of two main portions. In the first part of this function, from 0° to 45°, alpha influences the amount of backward masking while beyond 45° masking may be described as independent of orientation.

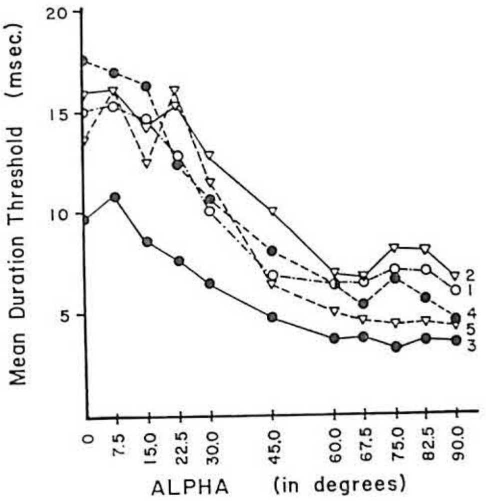


FIG. 1. Mean duration threshold as a function of alpha, the angle between test stripe, and masking contour orientation. (Test stripe oriented vertically; ISI = 40 msec. Each curve is for a single *O*, identified at right.)

The mean false-alarm rate, the probability of *O*'s responding "Yes" on catch trials, was 2% and was not related in any systematic fashion to alpha.

EXPERIMENT II

This was an attempt to extend the findings of the first experiment to the case in which the test stripe was oriented horizontally rather than vertically. Photometric and psychophysical conditions were as in Exp. I and the same *O*s served, the only difference being that a horizontal test stripe was used.

Results

The results of Exp. II are summarized in Fig. 2. For each of the five *O*s masking decreases with increasing alpha. This observation is confirmed by an analysis of variance. The effect of alpha was highly significant, F (11, 180) = 167.26, $p < .01$, while the $O \times$ Alpha interaction was nonsignificant ($F = 0.40$). In Exp. II, as in the first experiment, the two 0° alpha conditions did not differ significantly, t (44) = 0.66, and the point plotted against alpha = 0° is the mean for both conditions.

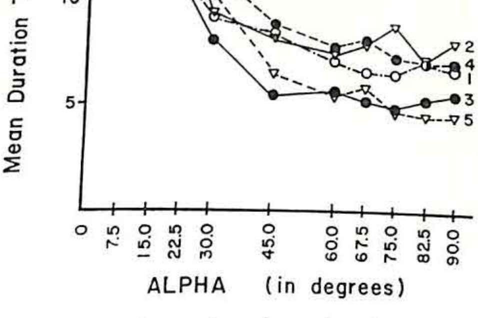


FIG. 2. Mean duration threshold as a function of alpha. (Test stripe oriented horizontally; ISI = 40 msec. Each curve is for a single *O* identified at right with same numbers as in Fig. 1.)

Tests of the significance of the difference between means of successive alphas indicated that beyond 60° the effect of increasing alpha was diminished below the point of statistical significance ($p > .05$). As in Exp. I the function relating backward masking to alpha seems to be composed of two parts, including both orientation-dependent and orientation-independent ranges. The mean rate of false-alarm responses was 1.6%.

The results of Exp. I and II indicate that for an ISI value of 40 msec. backward masking with both horizontal and vertical test stripes is a decreasing function of alpha. Experiment III was designed to measure the effect of alpha at several different ISIs, in order to assess the relationship between spatial and temporal determinants of backward masking.

EXPERIMENT III

Procedure

Three alphas (7.5, 22.5, and 60°) were studied in conjunction with ISIs of 5, 15, 30, 60, and 100 msec. Test-stripe orientation was horizontal. In any session *O* was tested at a single ISI with each of the three alphas presented in separate blocks of 120 trials. The *O* had been informed that the test stripe would be omitted on a randomly chosen .50 of all trials (actual rate of omission was .45–.55).

The tracking procedure, adapted from Campbell (1963), was run in blocks of eight trials with an intertrial interval of 5 sec. In the first block of eight trials the test stripe, when presented, was at 20 msec. If more than six responses to the eight trials had been correct the test stripe was presented at a reduced duration during the succeeding block. If fewer than six had been correct, duration was increased for the next block. If exactly six of the eight were correct, duration was unchanged. In this way *O*'s responses controlled the duration of the test-stripe presentations.

The *O* was given knowledge of results; a buzzer sounded coincidentally with each correct response. This feedback was provided in the belief that it might serve as additional incentive for *O*s operating under the increased strain of the reduced interval between trials. The order of presentation of ISI and alpha was randomized for each *O*. Three of the *O*s served in four sessions with each alpha and ISI combination while the remaining *O* served in only three sessions with each combination.

During the six practice sessions preceding the experiment measures of *O*'s normal, unmasked duration threshold were made. In addition, two series of 120 tracking trials each were run with neutral density filters substituted for the usual masking pattern. The luminance of this blank field was equivalent to the mean luminance of the usual masking pattern of alternating dark and bright stripes. One series was run with the homogeneous masking field at ISI = 5 msec., another at ISI = 100 msec.

Results

The results of Exp. III are summarized in Fig. 3. Each point is the mean taken over all *O*s and sessions with a particular combination of ISI and alpha. As might be expected from the results of earlier experiments (Kolers, 1962; Schiller, 1965; Schiller & Wiener, 1963), backward masking is a decreasing function of ISI. Both main effects, ISI and alpha, are very highly significant, $F\ (4, 12) = 108.71$; $F\ (2, 6) = 113.89$; $p < .001$, as is the interaction between them, $F\ (8, 24) = 9.02$. The difference between alpha $= 7.5°$ and $60°$ is highly significant ($p < .01$) for ISI = 5, 15, and 30 msec. ($t > 3.71$) while that between alpha = 7.5 and 22.5 is significant only at ISI = 15 msec. The pattern of differences among means for alpha-ISI combinations indicates that the effect of alpha upon backward masking may be greatest at the shorter ISI values.

The unmasked threshold of 3.77 msec. is indicated in Fig. 3 by the dashed line and lies below the duration threshold for any alpha-ISI combination. The location of the normal threshold suggests that some small amount of backward masking occurs even with ISI as great as 100 msec. The filled circles at ISI 5 and 100 msec. represent the mean duration thresholds obtained at those ISIs when a homogeneous masking field was used.

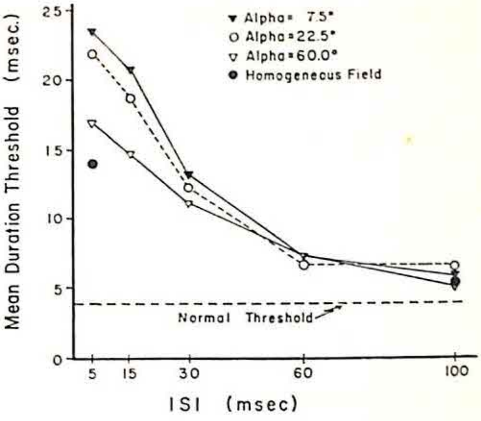


FIG. 3. Mean duration threshold as a function of alpha and ISI. (Test stripe oriented horizontally; four *O*s.)

Taking into account only the last six blocks of eight trials each, the mean false-alarm rate for each *O* was .22, .27, .27, and .18. There was no systematic relationship between false-alarm rate and either alpha or ISI.

DISCUSSION

Experiments I and II showed that backward masking can be related in a precise way to the relative orientations of test and masking contours. Masking decreased with increasing alpha. Experiment III analyzed the interaction between spatial and temporal determinants of backward masking. While the findings of the first two experiments were confirmed at the shorter ISI, at 60 and 100 msec. the effect of alpha had been substantially reduced. At these longer ISIs masking was not appreciably influenced by alpha.

The two 0° alpha conditions in each of the first two experiments differed in successive contrast between test stripe and the superimposed masking stripe. In one case, a bright masking stripe was superimposed upon the dark test stripe while in the other a dark masking stripe appeared in that position. Because of this successive contrast difference it was expected that the two conditions would yield dissimilar threshold values, though it is possible that the relatively low contrast of the stimuli used militated against such a finding.

It may be noted that none of the *O*s, when questioned, reported having seen any beta movement of test and masking stimuli, although such apparent movement has been observed in other masking studies (Fehrer & Smith, 1962).

The tracking procedures adapted for use in these experiments have proven to be useful and efficient. Cornsweet (1962) expressed concern that *O*s in studies employing a tracking technique would quickly become aware of the rules governing the stimulus and that this awareness would limit the value of tracking as a psychophysical method. Logan (1963) avoided this limitation, disrupting the tracking series by occasionally repeating the stimulus value from the immediately preceding trial without regard to *O*'s response. The present experiments used another procedure for disrupting the response-stimulus contingencies inherent in tracking. On some proportion of the trials the test stripe was omitted, this proportion varying from .28 in Exp. I and II to .50 in the last experiment. When questioned following their service, none of the *O*s reported that he had become aware of any relationship between his response and the stimulus duration on succeeding trials. In addition, as Swets (1961) has pointed out, the inclusion of many catch trials permits one to estimate the location of *O*'s response criterion, though this was incidental to the concerns of the present experiments.

REFERENCES

Battersby, W. H., & Wagman, I. H. Neural limitations of visual excitability. IV. Spatial determinants of retrochiasmal interaction. *Amer. J. Physiol.*, 1962, **203**, 359–365.

Battersby, W. S., & Wagman, I. H. Light adaptation kinetics: The influence of spatial factors. *Science*, 1964, **143**, 1029–1031.

Campbell, R. A. Detection of a noise signal of varying duration. *J. Acoust. Soc. Amer.*, 1963, **35**, 1732–1737.

Cornsweet, T. N. The staircase-method in psychophysics. *Amer. J. Psychol.*, 1962, **75**, 485–491.

Fehrer, E., & Smith, E. Effect of luminance ratio on masking. *Percept. mot. Skills*, 1962, **14**, 243–253.

Kinsbourne, M., & Warrington, E. K. The effect of an after-coming random pattern on the perception of brief visual stimuli. *Quart. J. exp. Psychol.*, 1962, **14**, 223–234. (a)

Kinsbourne, M., & Warrington, E. K. Further studies on the masking of brief visual stimuli by a random pattern. *Quart. J. exp. Psychol.*, 1962, **14**, 235–245. (b)

Kolers, P. A. Intensity and contour effects in visual masking. *Vis. Res.*, 1962, **2**, 277–294.

Kolers, P. A., & Rosner, B. On visual masking (metacontrast): Dichoptic observations. *Amer. J. Psychol.*, 1960, **73**, 2–21.

Logan, J. The application of the up-down method to psychophysical research. Paper read at Canadian Psychological Association, Quebec, June 1963.

Schiller, P. H. Monoptic and dichoptic visual masking by patterns and flashes. *J. exp. Psychol.*, 1965, **69**, 193–199.

Schiller, P. H., & Wiener, M. Monoptic and dichoptic visual masking. *J. exp. Psychol.*, 1963, **66**, 386–393.

Swets, J. A. Detection theory and psychophysics: A review. *Psychometrika*, 1961, **26**, 49–63.

(Received June 20, 1964)

Journal of Experimental Psychology
1965, 70, No. 4, 407–411

SIZE AND STRUCTURE OF GRAMMATICAL UNITS IN PAIRED-ASSOCIATE LEARNING AT TWO AGE LEVELS[1]

JAMES G. MARTIN AND ROBERT L. JONES

Chico State College

This study examined the functional equivalence of words and phrases and the effects of within-phrase redundancy in paired-associate learning with children and college students as Ss. There was no difference between single nouns and noun phrases as stimulus terms for both high- and low-redundant phrases. On the other hand, while high-redundant verb phrases were learned as quickly as single verbs as responses, low-redundant phrases were learned more slowly. Age did not interact with any of the grammatical variations.

Writers on the topic of language behavior have pointed out the inadequacies of associative-chain theories and Markov processes for explaining the construction of speech utterances. Lashley (1951), e.g., indicated that there is simply not enough time in the rapid execution of the motor skills involved in speech for control by some associative mechanism of stimulus-response chaining. Chomsky (1957) has shown that Markov processes do not predict many grammatical sentences which can be said and understood by users of the language. Miller, Galanter, and Pribram (1960) point out the impossibility of acquiring a language solely on the basis of past experience with even a large number of sentences.

Nevertheless the possibility cannot be ruled out that language facility begins with simple associations which become organized through experience into complex simultaneous and sequential hierarchies of behavior. Perhaps "novel and appropriate" sentences (Brown, 1958) consist not only of new sequences of words but also of novel combinations of whole phrases which have been overlearned and thus are easily available for sentence construction. These overlearned whole phrases may function as units of speech; through experience they become integrated (Osgood, 1957).

At the same time, these phrases may vary in degree of integration or redundancy as a function not only of experience but also internal structure. The word UP, for example, is more likely to follow WOKE than BEEN is to follow HAD, even though HAD BEEN, an auxiliary verb form, is probably more frequent in experience. But WOKE UP is more redundant because there are fewer probable alternatives for the second word.

These speculations were tested in the present experiment. First, two-word phrases in "sentences" learned by the paired-associates method were pitted against single words to determine the extent of their functional equivalence. Second, phrases differing in internal grammatical structure were compared to investigate

---

[1] Supported by Grant MH 08661-01, National Institute of Mental Health, United States Public Health Service. Appreciation is due to Ted DeBernardi, Principal, and the teachers of the Amer Jay Hamilton School for their cooperation in this research. A portion of this paper was reported at the American Psychological Association meetings, Los Angeles, 1964.

the relevance of a Markov interpretation to at least one level of sentence learning.

Since grammatical integrations develop with experience, grammatical relations between materials should facilitate learning more for adults than for children. To test this possibility both college and grade-school *S*s were used.

METHOD

Learning materials.—Ten basic four-word sentences each consisting of a noun phrase and a verb phrase were constructed. The noun phrases or parts thereof served as stimulus items in paired-associate learning, the verb phrases or parts thereof as response members. The semantic relations between noun and verb phrases were chosen to be unlikely. The sentences were constructed so that half of the noun phrases could be regarded as high redundant and half as low, and similarly for the verb phrases. Table 1 gives the sentences.

For convenience the top and bottom five sentences in Table 1 and the phrases therein are designated A and B, respectively. On the noun side, A phrases are either adjective-noun or proper noun combinations and were defined as high redundant, while B phrases are nouns preceded either by an article or an adjective indicating an indefinite quantity and were defined as low redundant. On the verb side, A phrases were low-redundant auxiliary-verb forms while B phrases were high-redundant verb-adverb or verb-object combinations.

Within the formal grammatical limits just specified the choice of (redundant) noun A phrases and verb B phrases was governed by the judgment that the words within the phrases were commonly experienced together in heard or written speech. To provide a behavioral check on the judgments, however, the first word of each noun and verb phrase in Table 1 was used as a stimulus word in a free-association test administered to 51 fifth and sixth graders and 53 college *S*s. The mean relative frequency with which the second word of a phrase was given as an associate to the first was as follows for grade-school *S*s: high-redundant noun A phrases, 28%; low-redundant noun B phrases, 0%; low-redundant verb A phrases, 5%; high-redundant verb B phrases, 16%. For college *S*s the corresponding percentages were 27, 5, 6, and 17. These results are consistent with the assumptions of the experiment and the judgments invoked in the selection of materials.

TABLE 1

MATERIALS

Within-Phrase Structure	Basic Sentences
A	ALARM CLOCK—CAN SEE BUSY BEE—DOES NOT YANKEE DOODLE—HAD BEEN NORTH POLE—WILL FLY MICKEY MOUSE—COULD BE
B	AN APPLE—WOKE UP SOME BREAD—ROLLS OVER THE CARS—TASTE GOOD A TREE—SWEEPS FLOOR THE BIRDS—WEAR CLOTHES

Design.—Each of the 10 basic sentences was dissected and arranged to yield five pairings which define the five phrase-unit conditions. These are best described by example and appear in Table 2. The comparisons of interest are as follows. If single nouns and noun phrases are functionally equivalent, pairs from Cond. II and III should be learned at the same rate, but faster than pairs from Cond. I in which the stimulus words are not integrated. The functional equivalence of single verbs and verb phrases is similarly tested by a comparison of III and IV, both of which should be learned faster than V.

TABLE 2

EXAMPLES OF EACH EXPERIMENTAL CONDITION

Within-Phrase Structure	Phrase-Unit Cond.				
	I	II	III	IV	V
A	CLOCK ALARM-CAN	ALARM CLOCK-CAN	CLOCK-CAN	CLOCK-CAN SEE	CLOCK-SEE CAN
B	APPLE AN-WOKE	AN APPLE-WOKE	APPLE-WOKE	APPLE-WOKE UP	APPLE-UP WOKE

The effects of redundancy within phrases are tested by a comparison of Cond. IIA and IIB on the noun side, and by IVA and IVB on the verb side.

Five mixed lists were constructed from the 10 basic sentences such that each phrase-unit condition was represented twice on each list, once for each within-structure condition. Thus each of the 10 experimental conditions was represented by a different sentence on a given list. Phrase-unit conditions and sentences were counterbalanced across lists so that each sentence appeared in each phrase condition on a different list. On each list the 10 pairs making a trial were randomized five times. Order of presentation was parallel across the five lists so the only difference between lists was the phrase condition in which a given sentence appeared.

Apparatus and procedure.—Lists were presented on a memory drum set behind a black screen. A light placed on the drum illuminated the aperture. The experimental room was dimly lit.

The *S*s were given standard instructions for paired-associate learning by the anticipation method. Both stimulus and response terms were to be pronounced. The necessity for repeating materials exactly as presented was emphasized. Pairs were presented at a 2:2-sec. rate with a 4-sec. interpair interval and a 16-sec. intertrial interval. A practice list of three nonsense pairs, GUZ VAP-WIK, ZUK-PIM GOL, and REG-DAT was learned to 1 perfect trial. Then the test list was presented for 10 trials, 1 study plus 9 anticipation.

Subjects.—There were 40 grade-school and 40 college *S*s, all with English as their native language. Half of the grade-school *S*s were fifth graders, half sixth graders, from the college laboratory school. The college *S*s were from introductory psychology classes and participated for extra credit or as a class requirement. Some had been in verbal-learning experiments before.

RESULTS

Table 3 presents mean errors over 10 trials according to phrase-unit condition, within-phrase structure, and age level. The statistical tests discussed below are based on an analysis of variance involving one between- and three within-*S*s error terms.

Age.—Table 3 shows that college *S*s learned faster over the whole list; the Age effect is significant, $F\ (1, 78) = 15.99$, $p < .0005$. However, the nonsignificant interactions of age level with anything else in the table indicate that grade-school *S*s react to the grammatical variations in this experiment in about the same way as college *S*s. It is clear that grammatical integrations are as effective at this developmental level as later so the age factor will not be considered further in the following analyses.

Units.—Table 3 shows mean errors over all *S*s for each of the phrase-

TABLE 3

MEAN ERRORS

Age Level	Within-Phrase Structure	Phrase-Unit Cond.					*M*
		I	II	III	IV	V	
Grade School	A	6.28	4.80	5.43	6.05	7.58	6.03
	B	5.35	4.48	4.38	4.28	6.25	4.95
	M	5.81	4.64	4.90	5.16	6.91	5.49
College	A	4.53	2.98	3.50	5.10	6.33	4.49
	B	4.15	3.10	3.58	3.48	4.48	3.75
	M	4.34	3.04	3.54	4.29	5.40	4.12
All *S*s	A	5.40	3.89	4.46	5.58	6.95	5.25
	B	4.75	3.79	3.98	3.88	5.36	4.35
	M	5.08	3.84	4.22	4.73	6.16	4.80

unit conditions. The differences were reliable, F (4, 312) = 24.31, $p < .0005$.

It was expected that pairs with integrated phrases on either the noun or verb side would be learned as fast as pairs with the corresponding single-word verbs and nouns. On the noun side first, the bottom line of Table 3 shows similar mean errors for noun phrases (Cond. II) and single-word nouns (Cond. III). The slightly lower error rate for phrases was not reliable ($p < .20$). But both single-word nouns and noun phrases were learned faster than the comparison condition (I), t (312) = 3.34 and 4.83, $p < .001$.

On the verb side, Cond. III phrases were learned faster than the comparison condition (V), t (312) = 7.57, $p < .001$. Phrases IV were also learned faster than V, t (312) = 5.56, $p < .001$, but not as fast as the single verbs (Cond. III), t (312) = 1.98, $p < .05$. All significant differences between phrase-unit conditions except the last were also significant ($p < .01$) by the Tukey test for multiple comparisons.

The results for noun phrases support the hypothesis that phrases may function like single words in paired-associate learning. Similar results did not hold over the verb phrases lumped together, but it can be seen in Table 3 that the difference between single-word verbs and verb phrases was a function of the structure of the verb phrases. The implication of this difference is discussed in the next section.

Within-phrase structure.—In the construction of basic sentences, phrases of differing structure were chosen presumably to vary in redundancy. It can be seen in Table 3 that the two noun-phrase structures in Cond. II differed only negligibly. None of the differences between Structures A and B in Cond. I, II, and III were significant. But the difference between verb phrases was substantial. Note that although pairs with high-redundant verb phrases (IVB) were learned as fast as pairs with single-word verbs (IIIB), the pairs with low-redundant auxiliary-verb combinations (IVA) fared badly compared to the corresponding single-verb pairs (IIIA). This difference was tested by the simple 2 × 2 interaction between these four cells; F (1, 312) = 5.05, $p < .05$. This finding shows that within-phrase structure can predict the availability of phrases in paired-associate learning.

DISCUSSION

This study has demonstrated the functional equivalence of some phrases and single words in a conventional learning task. It should be noted that pairs including phrases may be at a disadvantage in this task where they must be pronounced within the 2-sec. anticipation interval. Yet in general the performance with integrated phrases was equal to that with single words, and this fact suggests that in any theory of language the units of speech should not be limited to words.

Although learning rates differed between high- and low-redundant verb phrases as response terms, there was no difference with varying redundancy in the noun phrases. This outcome is probably an artifact of the paired-associate method, since *S* needed only to pronounce but not to recall the noun phrases as stimulus terms.

The fact that grammatically ordered combinations of words were more easily linked to their corresponding pairwise terms than backward combinations, especially when the order varied only on the stimulus side of the pairs (Cond. I and II), is of theoretical interest. Although the finding is consistent with several current interpretations of language, e.g., Jenkins and Palermo's (1964) hypothesis of sequences of linguistic classes, Miller's (1956) chunking hypothesis, and Osgood's (1957) integration-level hypothesis, it does not permit a clear choice of one over the alternative accounts. The theoretical interpretation of differences as a function of within-phrase redundancy, however, is clearer and seems to suggest a Markovian model for at least some levels of grammatical analysis.

REFERENCES

BROWN, R. W. *Words and things.* Glencoe, Ill.: Free Press, 1958.

CHOMSKY, N. *Syntactic structures.* 's-Gravenhage, Holland: Mouton, 1957.

JENKINS, J. J., & PALERMO, D. S. Mediation processes and the acquisition of linguistic structure. In Ursula Bellugi & R. W. Brown (Eds.), The acquisition of language. *Monogr. Soc. Res. Child Developm.*, 1964, 29(1, Whole No. 92).

LASHLEY, K. S. The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior.* New York: Wiley, 1961. Pp. 112–135.

MILLER, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.*, 1956, **63**, 81–97.

MILLER, G. A., GALANTER, E., & PRIBRAM, K. H. *Plans and the structure of behavior.* New York: Holt-Dryden, 1960.

OSGOOD, C. E. Motivational dynamics of language behavior. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1957.* Lincoln: Univer. Nebraska Press.

(Received June 22, 1964)

Journal of Experimental Psychology
1965, 70, No. 4, 412–416

SONIC AND ULTRASONIC EQUAL-LOUDNESS CONTOURS[1]

JOHN F. CORSO
State University of New York, Cortland

AND

MURRAY LEVINE
Aberdeen Proving Ground, Aberdeen, Maryland

Equal-loudness contours were established and compared for 18 Os tested by air conduction from 2,000 cps to 16,000 cps and for 30 Os tested by bone conduction from 2,000 cps to 94,000 cps. Contours were established for both modes of transmission at 0-, 10-, and 20-phon loudness levels. The results indicate that the curves for the 2 modes are essentially similar up to 14,000 cps, approximately; however, the extrapolated air-conduction curves appear to converge at 17,000 cps, while the bone-conduction curves converge at 85,000 cps. It is concluded that the loudness function for human hearing extends considerably higher in frequency than previously believed.

In the study of human hearing, the acoustic stimuli may be presented by either of two primary routes: air conduction or bone conduction. These two modes of transmission have been reported to involve similar physiological mechanisms (Békésy, 1932; Lowy, 1942; Tonndorf, 1962; Wever & Bray, 1936). Consequently, it has tacitly been assumed that the two systems should yield similar psychophysical functions. Corso and Levine (1963) have shown that the two systems do generate pitch-discrimination functions of similar form, but the bone-conduction mode is significantly more sensitive. Furthermore, recent studies (Corso & Oda, 1962; Deatherage, Jeffress, & Blodgett, 1954; Haeff & Knox, 1963; Pumphrey, 1950) have shown that the upper limit of audibility for bone conduction extends at least to 100,000 cps, while that for air conduction terminates in the region of 25,000 cps (Corso & Oda, 1961; Wever & Lawrence, 1954).

The primary purpose of the present study, therefore, was to compare the two modes of acoustic transmission by establishing a set of equal-loudness contours for both air and bone conduction from 2,000 cps to a point near the upper limit of audibility for each mode.

METHOD

Observers.—The *O*s were 48 male students, 18–20 yr., who volunteered from an introductory course in psychology. All *O*s were screened for normal hearing and freedom from otological abnormalities. All *O*s were administered the Loudness test of the Seashore Measures of Musical Talent and only those *O*s who obtained a score at the seventy-fifth percentile or higher on retesting were retained as *O*s from a larger pool of 93 volunteers. The 48 *O*s were divided randomly into two groups: 18 were tested by air conduction and 30 were tested by bone conduction.

Apparatus.—The bone-conduction system was designed and constructed to provide acoustic stimuli from 1,000 cps to 94,000 cps at a sound pressure level (SPL) at least 20 db. above each *O*'s threshold. The stimuli were generated by a bank of seven oscillators, one of which provided the desired standard tone. The output of the oscillators could be controlled manually or automatically and was fed into a dual-channel electronic switch containing a General Radio decade attenuator in each channel. The attenuated output passed through an Eico power amplifier to a barium-titanate crystal transducer (Corso, 1963b). The input voltage to the transducer was monitored by a Hewlett-Packard voltmeter. A lever system was used to apply the tip of the transducer to *O*'s head with a force of 200–300 gm. A photocell arrangement attached to acoustical muffs worn over both ears by each *O* permitted *E* to monitor *O*'s head position. The acoustic output of the transducer was calibrated a few millimeters from its conical tip which was placed in contact with the surface of a 3-ft.^{3} tank of water.

---

[1] This experiment was conducted at the Pennsylvania State University and was supported by Research Grant B-2006(RI) to the first author from the National Institute of Neurological Diseases and Blindness of the United States Public Health Service. The authors are grateful to Fujio Oda and John Searle for the instrumentation and calibration of the electronic testing systems.

The air-conduction system was similar to that for bone conduction, except that the output of the power amplifier passed through a Krohn-Hite filter and was presented to an Electrovoice, T350, horn-type speaker or to an Acoustic Research Transducer for frequencies below 5,000 cps. The system generated SPLs of at least 100 db. re 0.0002 dyne/sq cm from 1,000 to 16,000 cps and was calibrated with a Western Electric 640-AA condenser microphone. A special probe-tube microphone monitored the SPL at the entrance of the external auditory meatus (Corso & Oda, 1961). As in the bone-conduction system *O* was provided with the head-monitoring device and wore an acoustic muff over the nontested ear. All tests, both air and bone, were conducted in an anechoic chamber equipped with a two-way voice communication system and a closed-circuit television system.

Procedure.—For air conduction, the 18 *O*s were subdivided randomly into three groups of 6; for bone conduction, the 30 *O*s were subdivided randomly into five groups of 6. All groups were given two 1-hr. test sessions. In the practice sessions, both air- and bone-conduction thresholds were established at 1,000 cps, 2,000 cps, and at two other predetermined frequencies as shown in Table 1. All *O*s also made equal-loudness judgments in both the practice and test sessions, matching a given frequency to a 1,000-cps standard. In the first test hour, the judgments were made at a 20-phon loudness level; at the second, a 10-phon loudness level.

Equal-loudness contours were obtained by the method of limits as modified by Thorne (1934), with two categories of response. Two ascending and two descending series were presented in random order, after a descending practice trial. The standard stimulus was always 1,000 cps at the appropriate loudness level. Both the standard and variable stimuli were presented for 1.5 sec. with a .5-sec. separation. An interval of approximately 2 sec. was allowed for *O*'s response. The intensity of the comparison stimulus was changed in 1-db. steps and each series was terminated when *O* reported three consecutive judgments in the direction opposite that for the beginning trial.

RESULTS

In the analysis of data, the point of subjective equality (PSE) was determined for each test series and the mean value for the four series was taken as *O*'s PSE for a given frequency. The results for air conduction are presented graphically in Fig.

TABLE 1

DESIGN OF STUDY OF EQUAL-LOUDNESS CONTOURS

Cond.	Frequencies Tested (cps)				
	Group 1	Group 2	Group 3	Group 4	Group 5
Monaural Air Conduction (N = 18)	2,000 4,000 5,650	2,000 8,000 11,300	2,000 14,000 16,000		
Subgroup *n*	6	6	6		
Bone Conduction (N = 30)	2,000 4,000 5,650	2,000 8,000 11,300	2,000 14,000 16,000	2,000 57,000 64,000	2,000 85,000 94,000
Subgroup *n*	6	6	6	6	6

Note.—All groups were tested at threshold and at both 10- and 20-phon loudness levels.

1 which shows the contours without smoothing for the 0- (threshold), 10-, and 20-phon loudness levels.

In general, the curves of Fig. 1 resemble those in the literature (Fletcher & Munson, 1933; Robinson & Dadson, 1957). The curves begin to close in beyond 10,000 cps, with the extrapolated point of convergence at approximately 17,000 cps. This suggests that uniform changes in sound pressure from 0 to 20 phons produce fairly uniform (linear) changes in the underlying loudness function for frequencies from 1,000 cps to 10,000 cps, but that a nonlinear segment exists from 10,000 cps upward. It appears that the loudness function for air conduction terminates in the region of 17,000, i.e., near the upper limit of audibility.

The low threshold value at 8,000 cps should also be noted. This increase in sensitivity has been reported in earlier studies (Corso, 1958, 1963a). Since the present data are derived from a free-field technique rather than minimum audible pressure, it may be concluded that the 8,000-cps "notch" can no longer be attributed to earphone coupling effects.

The equal-loudness contours for bone conduction are shown in Fig. 2. Unlike the curves of Fig. 1, these curves are essentially parallel up to and including 16,000 cps and then converge at approximately 85,000 cps. A comparison of the curves in Fig. 1 and Fig. 2 suggests that the underlying loudness functions for air and bone conduction are not identical. While the two functions are probably

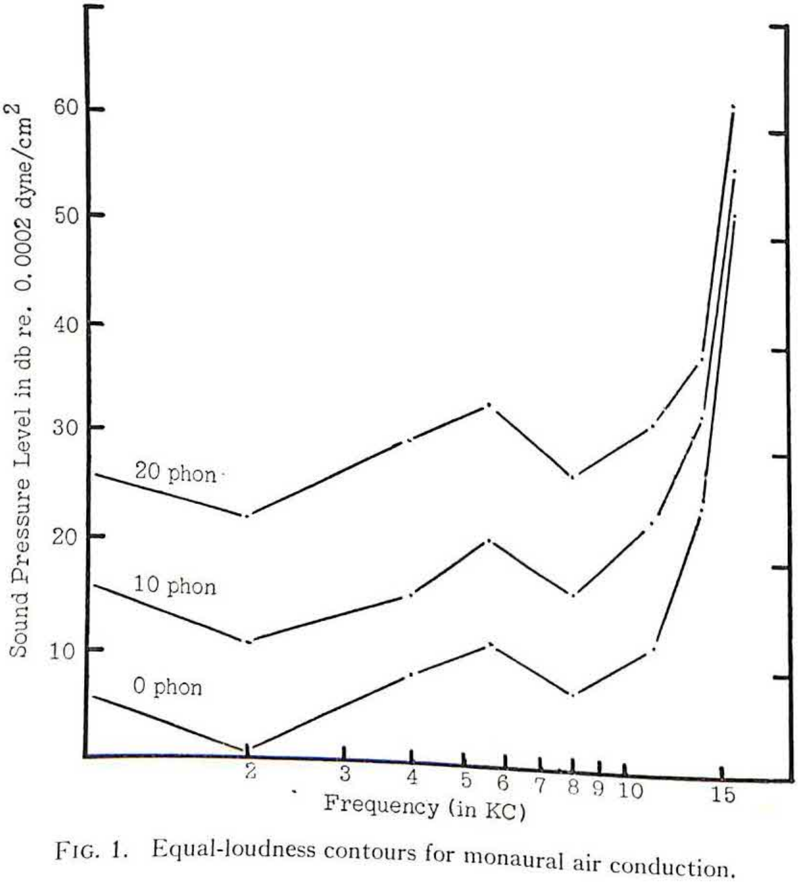


FIG. 1. Equal-loudness contours for monaural air conduction.

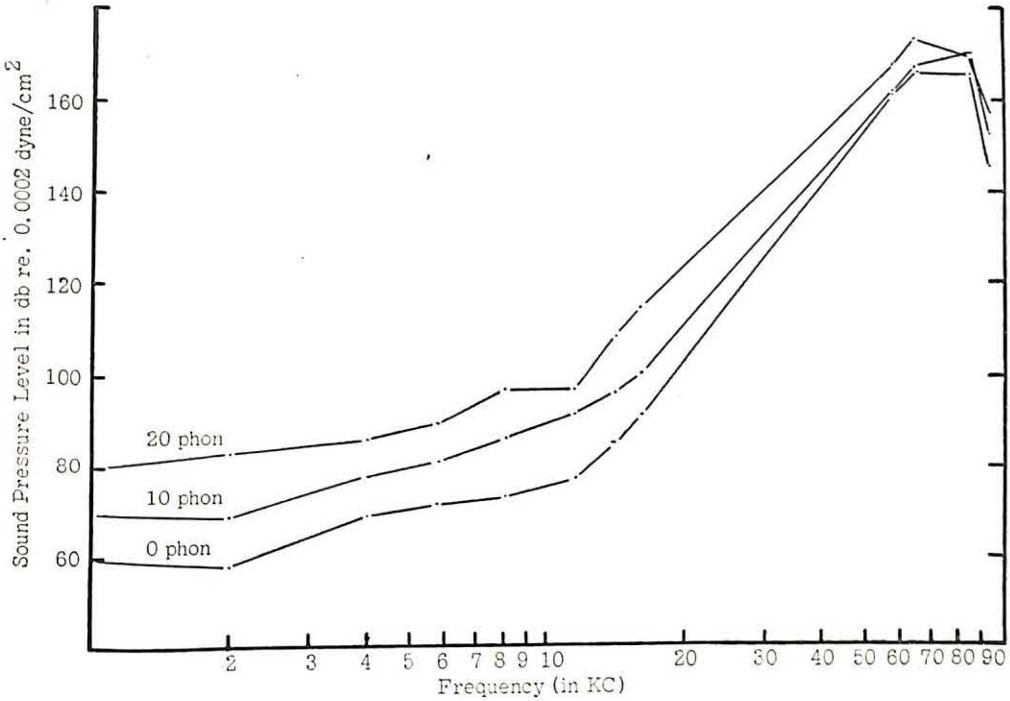


FIG. 2. Equal-loudness contours for bone conduction.

quite similar to approximately 14,000 cps, the loudness function for bone conduction must extend well beyond that for air conduction. It should also be observed that for bone conduction there is no increase in sensitivity at 8,000 cps, indicating that the air-conduction notch may be related to the functioning of a noncochlear element such as the tympanic membrane or the ossicular chain.

In Fig. 2 the surprising increase in sensitivity at 94,000 cps is an anomaly. Moreover, the bone-conduction curves separate again at this point for no apparent reason.

For both air and bone conduction, the standard error of the mean (*SE*) at each frequency is fairly large. The values range from 1.3 to 7.8 db. for air conduction and from 1.5 to 5.0 db. for bone conduction. In most cases, the *SE*s for the threshold determination (0 phons) are smaller than those for the equal-loudness matches at 10 and 20 phons.

DISCUSSION

The results of this study indicate that the equal-loudness contours for air and bone conduction are generally similar in form from 2,000 cps to approximately 14,000 cps. However, the air-conduction contours appear to converge at 17,000 cps, approximately, while the bone-conduction contours appear to converge at 85,000 cps, approximately.

These findings, when considered in conjunction with the results of a study on pitch discrimination (Corso & Levine, 1963), indicate that the auditory system is capable of processing loudness-discrimination information but not pitch-discrimination information for bone-conducted tones in the ultrasonic region from 20,000 cps to at least 85,000 cps. Pitch discrimination seems to fail near the upper limit of hearing by air conduction, but equal-loudness judgments may be made reliably for at least another four octaves. This suggests that the limit for pitch discrimination for both air and bone conduction is imposed by the neuroanatomical structure of the cochlea, whereas the limit for loudness discrimination in air conduction is imposed by some peripheral structure of the auditory system, such as the ossicular chain.

The available data indicate that the auditory mechanisms mediating pitch and loudness discrimination do not fail at the same upper frequency limit. When the limiting factor of the air-conduction system is by-passed by resorting to bone conduction, the range of frequencies over which the ear can respond in loudness is considerably greater than previously believed. Throughout this ultransonic range, when *O* is presented with a given frequency, he hears a tone whose pitch is equated to that of a bone-conducted sound at approximately 16,000 or 17,000 cps (Corso & Levine, 1963), but whose loudness may be made to vary. Caution should be exercised in interpreting the results of the present study on equal-loudness contours, however, since the signal levels in the ultrasonic region were relatively high and may have generated some extraneous cutaneous effects, thereby contaminating or supplanting pure loudness judgments. On the other hand, it should be noted that no *O*s gave any indication that the stimulus cues to which they were responding were other than auditory in nature.

REFERENCES

Békésy, G. von Zur Theorie des Hörens bei der Schallaufnahme durch Knochenleitung. [On the theory of audition in sound reception by bone conduction.] *Ann. d. Physik*, 1932, **13**, 111–136.

Corso, J. F. Proposed laboratory standard of normal hearing. *J. Acoust. Soc. Amer.*, 1958, **30**, 14–23.

Corso, J. F. Age and sex differences in pure-tone thresholds: Survey of hearing levels from 18 to 65 years. *Arch. Otolaryngol.* 1963, **77**, 385–405. (a)

Corso, J. F. Bone-conduction thresholds for sonic and ultrasonic frequencies. *J. Acoust. Soc. Amer.*, 1963, **35**, 1738–1743. (b)

Corso, J. F., & Levine, M. Pitch discrimination by air conduction and by bone conduction for sonic and ultrasonic frequencies. *J. Acoust. Soc. Amer.*, 1963, **35**, 804. (Abstract)

Corso, J. F., & Oda, F. Preliminary report on upper frequency audibility functions. Research Bulletin No. 17, 1961, Pennsylvania State University, Department of Psychology.

Corso, J. F., & Oda, F. Bone-conduction thresholds for sonic and ultrasonic frequencies. *J. Acoust. Soc. Amer.*, 1962, **34**, 746. (Abstract)

Deatherage, B. H., Jeffress, L. A., & Blodgett, H. C. A note on the audibility of intense ultransonic sound. *J. Acoust. Soc. Amer.*, 1954, **16**, 582.

Fletcher, H., & Munson, W. A. Loudness, its definition, measurement and calculation. *J. Acoust. Soc. Amer.*, 1933, **5**, 82–108.

Haeff, A. V., & Knox, C. Perception of ultrasound. *Science*, 1963, **139**, 590–592.

Lowy, K. Cancellation of the electrical cochlear response with air and bone-conducted sound. *J. Acoust. Soc. Amer.*, 1942, **13**, 156–158.

Pumphrey, R. J. Upper limit of frequency for human hearing. *Nature*, 1950, **166**, 57.

Robinson, D. W., & Dadson, R. S. Threshold for hearing and equal loudness relations for pure tones, and the loudness function. *J. Acoust. Soc. Amer.*, 1957, **29**, 1284–1288.

Thorne, F. C. The psychophysical measurement of the temporal course of visual sensitivity. *Arch. Psychol., N. Y.*, 1934, No. 1700.

Tonndorf, J. Compressional bone conduction in cochlear models. *J. Acoust. Soc. Amer.*, 1962, **34**, 1127–1131.

Wever, E. G., & Bray, C. W. The nature of bone conduction as shown in the electrical response of the cochlea. *Ann. Otol. Rhinol. Laryngol.*, 1936, **45**, 822–830.

Wever, E. G., & Lawrence, M. E. *Physiological acoustics.* Princeton: Princeton Univer. Press, 1954.

(Received June 22, 1964)

Journal of Experimental Psychology
1965, 70, No. 4, 417–423

A REVERSED PARTIAL-REINFORCEMENT EFFECT[1]

WILLIAM B. PAVLIK AND PETER L. CARLTON

Rutgers University

The effects of partial- and continuous-reinforcement schedules were compared on both a between- and a within-Ss basis. Rats given either partial or continuous reinforcement in a lever-pressing situation (between-Ss comparison) showed the usual partial-reinforcement effects; both greater resistance to extinction and higher terminal acquisition response rates were associated with partial reinforcement. The same pattern of results did not, however, appear when a within-Ss comparison was made. Rats exposed to both continuous- and partial-reinforcement schedules showed higher terminal acquisition performance in the partial-reinforcement condition, but showed greater resistance to extinction in the continuous-reinforcement condition.

Grice and Hunter (1964) have recently pointed out that between-*S*s comparisons (i.e., each *S* is exposed to but a single experimental condition) yield stimulus intensity effects that differ from those obtained from within-*S*s comparisons (i.e., each *S* is exposed to more than one experimental condition). That a similar difference may occur in the study of the effects of reinforcement schedules in the instrumental-reward situation is suggested by a search of the literature and the results of the present study.

Between-*S*s comparisons typically have shown that both higher levels of terminal acquisition performance and greater resistance to extinction are associated with schedules having lower densities (smaller percentages) of reinforcement (e.g., Goodrich, 1959; Wagner, 1961; Weinstock, 1958). Within-*S*s comparisons, on the other hand, frequently have shown higher levels of acquisition performance and greater resistance to extinction to be associated with higher density schedules (e.g., Clayton & Stevenson, 1961; Gollub & Urban, 1958). That such a difference in results is due to differences in experimental design is uncertain, however, as the studies involved typically have differed in other important respects. Most between-*S*s comparisons have involved discrete-trial situations and have compared continuous reinforcement (CRF) groups with fairly high-density partial-reinforcement (PR) groups. Most within-*S*s comparisons, on the other hand, have involved free-operant situations and have used relatively low-density schedules. An exception to the latter generalization is the recent runway study by Amsel, MacKinnon, Rashotte, and Surridge (1964), who showed, in a within-*S*s comparison, that higher terminal running speeds occurred in a stimulus condition associated with PR than in a different stimulus condition associated with CRF. None of these studies has, however, simultaneously used both experimental designs; therefore, direct comparisons of the effects yielded by the two designs are not possible.

Studies which have contrasted these experimental designs in a reinforcement-schedule context have been reported by D'Amato, Lachman, and Kivy (1958) and by Pavlik, Carlton, and Hughes (1964). The former study showed, in a secondary-reinforcement situation, that the relationship of the strength of a secondary reinforcer to the training schedule of primary reinforcements depended upon the experimental design. The CRF gave stronger secondary reinforcement on the basis of a within-*S*s comparison, but PR gave stronger secondary reinforcement with a between-*S*s comparison. Pavlik et al. report a similar pattern of effects in a runway situation; resistance to extinction was greater following PR in the between-*S*s comparison, but tended to be greater following CRF in the within-*S*s comparison.

---

[1] This research was supported by United States Public Health Service Grant MH 07844.

The present study was designed to shed additional light on this problem by comparing the effects of CRF and PR reinforcement schedules in both between-*S*s and within-*S*s experimental designs. Although a free-operant situation was used, a relatively high-density PR schedule (VR-3) was employed so as to enhance comparability to the typical between-*S*s study.

METHOD

Subjects.—The *S*s were 24 naive male rats, 180–200 gm. at the start of the experiment, which were purchased from Sprague-Dawley, Incorporated.

Apparatus.—A standard two-lever Grason-Stadler response chamber was used. The sound-insulated chest housing the chamber contained a Gerbrands pellet dispenser that could deliver 45-mg. Noyes pellets into the food cup mounted in the response chamber. Attendant automatic programing equipment was used to instrument the reinforcement contingencies to be described.

The houselight ordinarily used with this apparatus was disconnected. Illumination was provided by either the white cue light mounted over the right-hand lever or the red cue light mounted over the left-hand lever. Such illumination defined the discriminative stimuli involved in the within-*S*s comparisons described below. Throughout the experimental sessions, white noise (from a Grason-Stadler noise generator set at −15 db.) was delivered to the speaker mounted on the response chamber.

Procedure.—The experimental procedure can conveniently be divided into three phases. In Phase 1 of the experiment, *S*s were placed on a deprivation regimen under which they received a daily ration of 10 gm. of standard Purina lab chow. On those days on which *S* received food in the response chamber, the home-cage ration was adjusted so that the total amount of food (Noyes pellets plus lab chow) came to 10 gm. The home-cage ration was administered 30–60 min. after the daily experimental sessions.

On the fourth day following the beginning of the deprivation regimen, Day 4, each *S* was placed in the response chamber for 30 min. During this time, pellets were delivered automatically at a rate of 2/min. On Day 5, each *S* again was placed in the response chamber for 30 min. During this period, pellets could be obtained only by *S*'s pressing the right-hand lever. Each lever press delivered one pellet (CRF). Day 6 was identical with Day 5 except that the session was terminated when *S* had received a predetermined number of pellets. This number was adjusted for each *S* so that the total number of pellets received on Days 5 and 6 was 250 for each *S*.

Phase 2 began on Day 7. The *S*s were randomly assigned to four groups of 6 *S*s each. For Group 1, depressions of the right-hand lever led to the delivery of pellets on a CRF schedule. The *S*s in Group 2 received pellets on a PR schedule such that an irregular one third of lever presses led to the delivery of a pellet. Thus, Groups 1 and 2 provide for conventional between-*S*s comparisons of the effects of the schedules.

Group 3 and Group 4 *S*s were exposed to both reinforcement schedules, and thus provided within-*S*s comparisons. For three of the *S*s in each of these groups, illumination of the white cue light was associated with CRF and the red cue light with the PR schedule. The reverse correlation of cue light with schedule obtained for the other three *S*s in each of these groups.

Experimental sessions in Phase 2 were of 10-min. duration for *S*s in Groups 1 and 2, and of 20-min. duration for *S*s in Groups 3 and 4. The 20-min. sessions for Groups 3 and 4 involved 10-min. exposure to each of the reinforcement schedules, thus permitting direct comparisons to be made with Groups 1 and 2. All groups were given such sessions on Days 7–10.

Phase 3 took place on Day 11 and was identical in all respects to Phase 2 except that no pellets were delivered. During this single extinction session, the pellet hopper was empty; the magazine continued to operate as usual. Groups 1 and 2 received 10 min. of extinction; Groups 3 and 4 received 20 min. of extinction, 10 min. with the cue light associated with CRF and 10 min. with the cue light associated with PR. This again provided for direct comparisons of Groups 3 and 4 with Groups 1 and 2.

Throughout *all* three phases of the experiment for all *S*s, the red and white cue lights were alternately illuminated. For Groups 1, 2, and 3 the white light was on for the first 30 sec., the red for the next 30 sec., the white for the next 30 sec., etc. For Group 4 the red light was on for the first 30 sec., the white for the next 30 sec., etc. The *S*s in Group 4 were given their experimental sessions immediately following the completion of the sessions for Groups 1, 2, and 3. Although both cue lights were used, it should be emphasized that only the right-hand lever operated the magazine. The left-hand lever was nonfunctional throughout the experiment.

RESULTS

To provide for similar statistical analyses of between- and within-*S*s effects, *S*s in Groups 1 and 2 were paired on the basis of their response rates during the last session of Phase 1, prior to the introduction of the difference in schedules. Subsequent analyses were of difference scores. Between-*S*s comparisons were made by subtracting the score for each CRF *S* in Group 1 from the analogous score of his paired PR *S* in Group 2. Within-*S*s comparisons were made by subtracting the score for each *S* in Groups 3 and 4 in the CRF condition from his score in the analogous PR condition.

Terminal acquisition performance.—The numbers of responses emitted in the last session of Phase 2 were taken to represent terminal acquisition performance. Raw score means and *SD*s for the various groups are presented in Table 1. (Groups 3 and 4 have been combined as they differed only with respect to which cue light started the experimental sessions.) In the between-*S*s comparison, each Group 2 *S* responding under the PR schedule emitted more responses than did his paired CRF *S* in Group 1. The mean difference (148.8) was reliably different from zero ($p = .02$ by Wilcoxon's T test). As far as the between-*S*s comparison is concerned then, PR led to higher terminal acquisition performance than did CRF.

TABLE 1

RAW SCORE MEANS AND *SD*s FOR TERMINAL ACQUISITION AND EXTINCTION PERFORMANCE

Group	Acquisition		Extinction	
	M	*SD*	*M*	*SD*
1 (CRF)	110.7	11.9	65.8	15.8
2 (PR)	259.5	40.1	119.0	38.1
3 & 4 (CRF)	119.9	16.5	75.8	25.7
3 & 4 (PR)	163.4	33.7	51.8	20.3

Comparable results were obtained from the within-*S*s comparison. As indicated by the means in Table 1, *S*s in Groups 3 and 4 tended to respond more in the PR condition than in the CRF condition. The mean of the difference scores (43.5) was reliably different from zero ($p = .02$ by Wilcoxon's T test). Although this difference was reliable, it was not uniform for all *S*s. Four of the 12 *S*s in Groups 3 and 4 emitted more responses in the CRF condition than in the PR condition. These four inversions were, however, the smallest difference scores obtained.

A comparison of the mean acquisition scores in Table 1 indicates that the number of responses emitted by Group 1 ($M = 110.7$) did not differ appreciably from the number emitted by Groups 3 and 4 *S*s in the stimulus condition associated with CRF ($M = 119.9$). These means do not differ when evaluated by the U test. In contrast, Group 2 *S*s emitted more responses ($M = 259.5$) than did Groups 3 and 4 *S*s in the stimulus condition associated with PR ($M = 163.4$). These means differ reliably ($p < .001$ by the U test).

Thus, PR led to reliably higher rates of responding during acquisition in both the between- and within-*S*s comparisons. Further, *S*s exposed to both PR and CRF responded, during the CRF condition, in a manner comparable to *S*s exposed only to CRF for a similar time period; however, they emitted fewer responses during the PR condition than did *S*s exposed only to PR for the same length of time.

Extinction performance.—The data obtained during the extinction session are summarized in Fig. 1. Each bar in the left-hand portion of the figure, obtained by subtracting the number of responses emitted by an *S* in Group 1 from the number emitted by its paired *S* in Group 2, provides an estimate of the usual between-*S*s partial-reinforcement effect (PRE). In each case, a greater number of responses was emitted in extinction by the *S* trained with PR ($p = .02$ by Wilcoxon's test). (Group means and *SD*s are given in Table 1.)

As the values in the right-hand portion of Fig. 1 indicate, the opposite effect, a reversed PRE, was obtained from the within-*S*s comparison. These values were obtained by subtracting the number of responses emitted in the CRF stimulus condition from the number emitted in the PR condition. Each bar represents the data from one *S*. With one exception, more extinction responses were emitted in the CRF stimulus condition than in the PR stimulus condition ($p < .01$ by Wilcoxon's test). (The single inversion amounts to one more response in the PR than in the CRF condition.) Means and *SD*s for the CRF and PR conditions are given in Table 1.

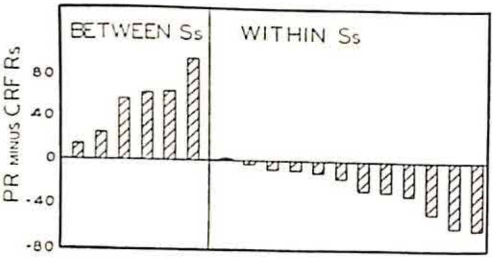


FIG. 1. Differences in resistance to extinction for partial vs. continuous reinforcement conditions. (The left portion represents between-*S*s comparisons, each bar depicting the difference between a single matched pair of *S*s. The right portion represents within-*S*s comparisons, each bar depicting the difference between schedules for a single *S*.)

It should be pointed out that the reversed PRE obtained with the within-*S*s comparison was affected, to some extent, by the particular combination of cue-light arrangements used. The three *S*s who had the white cue light as both the CRF condition and the light that started each session yielded much larger reversed PREs than did any of the *S*s in the other three combinations of CRF stimulus and starting stimulus. These other combinations appear not to differ.

Although the extinction responding of Group 1 ($M = 65.8$) tended to be less than that of Groups 3 and 4 *S*s in the CRF condition ($M = 75.8$), this difference did not attain significance by the U test. However, extinction responding by Groups 3 and 4 *S*s in the PR condition ($M = 51.8$) was reliably less than that of Group 2 *S*s ($M = 119.0$), $p = .002$ by the U test.

DISCUSSION

The usual between-*S*s PREs were obtained in this study; both terminal acquisition response rates and responses in extinction were greater with PR than with CRF schedules. Divergent results were obtained when these comparisons were made on a within-*S*s basis. Terminal acquisition response rates were greater, but responses in extinction were *fewer* under the PR condition.

As mentioned above, the reversed PRE in extinction for the within-*S*s comparison was affected by the combination of cue lights used. Those *S*s for whom the white cue light was both the starting color and the CRF condition yielded larger reversed PREs than did *S*s in any other cue-light combination. There would appear to be two factors operating to produce this interaction. First, since it is reasonable to assume an overall decline in response rate during the extinction session, regardless of stimulus condition, that stimulus that started the session would tend to be correlated with the greater number of responses, regardless of schedule. When the starting stimulus was also the CRF stimulus, a greater reversed PRE should be expected. If this were the only factor operating, however, then *S*s for whom the red cue light was both the starting and CRF stimulus should show differences as large as those shown by *S*s for whom white was the starting and CRF stimulus. This was not the case. It seems likely that a second factor affecting our results is based upon the facts that (*a*) the right-hand lever was the one on which all reinforcements were scheduled, and (*b*) the white cue light was mounted directly above the right-hand lever. This spatial correspondence of which cue light and relevant lever, and the fact that the Grason-Stadler equipment is prewired so that the cue light blinks with the delivery of a reinforcement, would tend to maximize the secondary reinforcing properties of the white cue light relative to those of the red cue light, which is mounted several inches from the right-hand lever. The enhanced secondary reinforcing properties of the white cue light should produce a larger reversed PRE when white was the CRF stimulus.

It was also found that terminal acquisition response rates under PR were less in *S*s exposed to both PR and CRF schedules (Groups 3 and 4) than in *S*s exposed only to the PR schedule (Group 2). This could be due to the fact that Groups 3 and 4 *S*s were given 20-, rather than 10-min. sessions. Any within-sessions decline due, for example, to progressive satiation, would tend to reduce response rate during the session. Unfortunately, our recording procedure did not permit a direct assessment of within-session rate changes. In any event, the lack of a comparable rate difference under the CRF conditions (Group 1 vs. Groups 3 and 4 CRF responses) suggests that PR responding is more susceptible to whatever within-sessions declines may have been operating than is CRF responding. A greater susceptibility to satiation of responding maintained by lower densities of reinforcement has been reported elsewhere (Carlton, 1961; Patrikiou & Keehn, 1964).

The above interpretation does not, however, account for the similar pattern of results obtained in extinction. Again, Groups 3 and 4 *S*s showed a depression of response rate in the PR condition when compared with Group 2, but a slight, insignificant enhancement of performance in the CRF condition when compared with Group 1. Obviously, progressive satiation effects could not have been operating during extinction.

This pattern of extinction results indicates that the reversed PRE obtained from the within-*S*s comparison was largely due to a depression of performance in the PR condition without a comparable depression (in fact, a slight enhancement) of performance in the CRF condition. A similar pattern of results has been obtained in other situations (Clayton & Stevenson, 1961; Gollub & Urban, 1958). In both of these studies each animal was trained to respond in two stimulus conditions, each stimulus being associated with a different reinforcement density. Reinforcements were then discontinued in both conditions. The results of both of these within-*S*s evaluations of the effects of prior reinforcement density on extinction are in agreement with ours; greater numbers of responses were emitted in the stimulus associated with the higher density of reinforcement. Unfortunately, these experiments did not include animals that were exposed to only *one* of the reinforcement schedules. Direct comparisons of the effects of within vs. between effects are therefore not possible. Accordingly, one cannot determine whether the divergence in extinction responding of animals exposed to both reinforcement schedules was due to a depression of responding in the stimulus associated with a low reinforcement density, an augmentation of responding in the high density component, or a combination of these.

The general nature of our data immediately suggests that a "contrast" effect is involved (cf., Reynolds, 1961, 1963). Studies of contrast typically involve training animals to respond to different stimuli, each associated with a different density of reinforcement. Reinforcement density in one of these stimuli is then varied while the density in the second is held constant. "Contrast" is reflected by changes in responding in the second component that are opposite to the changes produced in the first component. Although our procedures are not identical to those operations that typically define "contrast," it could be argued that the depression of responding in the PR component, in both acquisition and extinction, is in some way due to a "contrast" with the CRF component. The magnitude of the "contrast" could presumably be measured in terms of the deviation of the PR and CRF responding of the animals exposed to both schedules from the corresponding performance of the animals exposed to only one of the schedules.

An interpretation in terms of "contrast" leads to the expectation that the acquisition PR responding of the animals exposed to the two schedules would be depressed with respect to those exposed only to PR; this expectation is borne out. Similar considerations lead to the expectation that, for the animals given both CRF and PR, their CRF responding should be relatively inflated. There was, however, only a nonsignificant suggestion of just such an effect. The failure to find a significant effect on CRF responding could have been due to a "ceiling" on the CRF responding in acquisition; the animals may simply have been unable to lever press and eat the subsequently delivered pellet at an appreciably higher rate. In extinction, it might also be expected that, in the within-*S*s comparison, CRF responding would be inflated and PR responding depressed relative to the between-*S*s comparison. Only the second of these results was obtained. The limitation of a "ceiling" on CRF behavior would not, obviously, apply. Reynolds (1963) has found, however, that "contrast" effects are minimal when the schedule to be "contrasted" is characterized by a high density of reinforcement. The minimal effects on responding in a stimulus previously associated with CRF may, therefore, reflect a carry-over of attenuated "contrast" into extinction.

It should be pointed out that there are few, if any, explanatory advantages in merely labeling our results as an instance of contrast. Because contrast, although an important phenomenon, is not well-understood, the simple categorization of our data as contrast adds little to the analysis.

A somewhat different interpretation of our data is possible. The main effect we have obtained is a depression of responding, in both acquisition and extinction, to the stimulus associated with PR. One possibility is that, in this stimulus, animals "wait" for the higher density reinforcement contingency scheduled in the CRF stimulus. That animals can, in fact, anticipate impending changes in reinforcement schedule is clearly illustrated in two studies by Pliskoff (1961, 1963). This interpretation is made especially tenable by the fact that the intervals of exposure were as short as 30 sec.

This interpretation does not account,

of course, for the mechanisms that control "waiting" by the within-*S* animals in the PR stimulus. It does, however, lead to a simple prediction. If the contingencies involved in switching from one stimulus to the other (from CRF to PR and vice-versa) are time based, then the animals can simply "wait" for a high-density schedule to be instated. This effect would be maximized by the short stimulus periods we have used. In contrast, the depression of responding in the PR stimulus should *not* be obtained when the switch from one stimulus to another is *response* based rather than time based. An experiment now in progress indicates that this is indeed the case.

One final point warrants comment. From a purely biological point of view, the within-*S*s variation in reinforcement contingencies seems to represent a substantially more probable state of affairs. It is difficult to conceive of an animal that, in a natural habitat, is exposed to only *one* reinforcement contingency. And, in this context, it is important to bear in mind that organisms, laboratory and otherwise, have evolved in such an environment. From this purely biological point of view, our results are intuitively reasonable; organisms should extinguish more slowly under circumstances in which the probability of reinforcement is highest. Also from this biological point of view, the typical between-*S*s increase in responding in a stimulus associated with PR may very well represent an interesting and important laboratory curiosity having very little relevance to the exigencies with which the animal, in its nautral environment, must cope.

REFERENCES

AMSEL, A., MACKINNON, J. R., RASHOTTE, M. E., & SURRIDGE, C. T. Partial reinforcement (acquisition) effects within subjects. *J. exp. Anal. Behav.*, 1964, **7**, 135–138.

CARLTON, P. L. The interacting effects of deprivation and reinforcement schedule. *J. exp. Anal. Behav.*, 1961, **4**, 379–381.

CLAYTON, F. L., & STEVENSON, J. G. Extinction of bar pressing following training under a multiple VI VI schedule. *J. exp. Anal. Behav.*, 1961, **4**, 295–297.

D'AMATO, M. R., LACHMAN, R., & KIVY, P. Secondary reinforcement as affected by reward schedule and the testing situation. *J. comp. physiol. Psychol.*, 1958, **51**, 737–741.

GOLLUB, L. R., & URBAN, J. T. The accentuation of a rate difference during extinction. *J. exp. Anal. Behav.*, 1958, **1**, 365–370.

GOODRICH, K. P. Performance in different segments of an instrumental response chain as a function of reinforcement schedule. *J. exp. Psychol.*, 1959, **57**, 57–63.

GRICE, G. R., & HUNTER, J. J. Stimulus intensity effects depend upon the type of experimental design. *Psychol. Rev.*, 1964, **71**, 247–256.

PATRIKIOU, V. D., & KEEHN, J. D. The effect of partial satiation on pauses after reinforcements scheduled on alternating fixed ratios. *Psychon. Sci.*, 1964, **1**, 79–80.

PAVLIK, W. B., CARLTON, P. L., & HUGHES, R. A. A comparison of between subjects with within subjects partial reinforcement effects. Paper read at Psychonomic Society, Niagara Falls, 1964.

PLISKOFF, S. Rate-change effects during a preschedule-change stimulus. *J. exp. Anal. Behav.*, 1961, **4**, 383–386.

PLISKOFF, S. Rate-change effects with equal potential reinforcement during the "warning" stimulus. *J. exp. Anal. Behav.*, 1963, **6**, 557–562.

REYNOLDS, G. S. Behavioral contrast. *J. exp. Anal. Behav.*, 1961, **4**, 57–72.

REYNOLDS, G. S. Some limitations on behavioral contrast and induction during successive discrimination. *J. exp. Anal. Behav.*, 1963, **6**, 131–139.

WAGNER, A. R. Effects of amount and percentage of reinforcement and number of acquisition trials on conditioning and extinction. *J. exp. Psychol.*, 1961, **62**, 234–242.

WEINSTOCK, S. Acquisition and extinction of a partially reinforced running response at a 24-hour intertrial interval. *J. exp. Psychol.*, 1958, **56**, 151–158.

(Received June 25, 1964)

Journal of Experimental Psychology
1965, 70, No. 4, 424–427

DISCRIMINATION SHIFTS AS A FUNCTION OF DEGREE OF TRAINING IN CHILDREN [1]

JAMES YOUNISS AND HANS G. FURTH

Center for Research in Thinking and Language, Catholic University of America

Children initially performed a 2-choice discrimination with variable irrelevant cues, either to criterion (CR) or with overtraining (OT) of 15 and 25 trials. Subsequently ½ of the Ss performed a reversal shift and ½ an extradimensional nonreversal shift. No shift difference was observed after CR; however, after OT, reversal, but not nonreversal shift, was facilitated with rapid acquisition of reversed responses. Data were accounted for by a dimension-specific mediating response which became stronger and a change in reinforcement which became more discriminable following OT.

Investigators have consistently observed that college students made reversal shift with less difficulty than extradimensional nonreversal shift (cf. Kendler & Kendler, 1962) while rats and young children found reversal to be the more difficult shift (Kelleher, 1956; Kendler, Kendler, & Wells, 1960). In addition to replicating the greater difficulty for reversal shift in rats, some investigators (Brookshire, Warren, & Ball, 1961; Mackintosh, 1962) reported that overtraining facilitated performance of reversal but not of nonreversal shift.

These latter observations suggest the interesting possibility that relative shift difficulty might interact with degree of training through a selective facilitation of reversal shift due to overtraining. Further suggestive evidence for such an expectation comes from a recent study with children (Furth & Youniss, 1964) which found that a reversal-type of intradimensional shift was made easier, while an extradimensional shift was little affected, by overtraining. The present study was designed to test directly for the effect of overtraining on type of shift in children within a previously employed reversal-nonreversal shift paradigm.

METHOD

Design.—A 3 × 2 × 2 factorial design included three amounts of postcriterion overtraining—0, 15, or 25 trials—and two shifts—either reversal (R) shift or extradimensional nonreversal (NR) shift. Additionally one half the *S*s were trained on a form-relevant discrimination while the other half were trained on a color-relevant problem.

Subjects.—From the first grade of a public elementary school 105 children were sampled, mean age = 77.12 mo., *SD* = 5.31; range 69–96 mo. Of these *S*s, 9 were eliminated for failure to learn the original discrimination.

Material.—Stimuli were 24 wood cuts of six colors: red, yellow, white, green, blue, and black, and four forms: circle, ellipse, triangle, and square. The forms were constructed so that the circle's diameter (3 in.) was equal to the square's sides, ellipse's length, and triangle's height.

The presentation tray, 14 × 8 in., was made of plywood and had two reward wells separated 8 in., center to center. A Masonite shield could be inserted in grooves in front of the wells between trials or behind the wells during trials.

Procedure.—Individual testing, carried out in a quiet room during class hours, began with instructions that *S* was to find a token under one of the blocks each time they were shown and that when enough tokens were won he could exchange them for a nickel prize.

---

[1] This research was supported by National Institute of Mental Health Grant 08614-01. Thanks are expressed to the administrative personnel of Prince Georges County, Maryland, who provided their cooperation.

TABLE 1

Means and *SD*s of Transformed Trials to Criterion in Training

Relevant Dimension	Shift	Degree of Training		
		0 Trials	15 Trials	25 Trials
Color	Reversal	1.96 (1.22)	2.63 (1.39)	2.18 (1.19)
Form	Reversal	2.62 (1.06)	2.87 (1.07)	2.20 (0.96)
Color	Nonreversal	1.82 (1.23)	2.22 (0.90)	2.45 (0.75)
Form	Nonreversal	1.97 (1.56)	1.90 (1.20)	2.39 (1.48)

A variable-irrelevant *training procedure* (House & Zeaman, 1962) with noncorrection was used. Each *S* was assigned in order of appearance to an experimental condition ($N = 8$). Further, *S* was assigned randomly a pair of cues from each of the color and form dimensions with the restriction that neither black and blue nor blue and green were paired. From one dimension a pair of cues was made relevant; i.e., one member of this pair was associated with reward on every trial while the other member was not rewarded. A pair of cues from the other dimension was uncorrelated with reward since each member would appear equally often with each member of the relevant pair over trials. A predetermined order of stimulus presentation for 30 trials counterbalanced position of the four cues.

If *S* failed to reach a criterion of 9 correct responses in 10 consecutive trials within a maximum of 30 trials he was eliminated from the study. For those *S*s who attained the criterion, 0, 15, or 25 *overtraining trials* were administered as a continuation of training.

Noncorrected *shift trials* began immediately after overtraining or criterion. Variable-irrelevant cue conditions prevailed on these trials also (House & Zeaman, 1962). For R shift that pair of cues which was relevant during training remained relevant but reward was switched from one to the other cue member which was previously unrewarded. Additionally two new cues from the previously irrelevant dimension were introduced in lieu of the previously irrelevant pair and these cues were also irrelevant. Thus, when color was relevant in training with red rewarded vs. blue unrewarded and circle and triangle were irrelevant, blue was reinforced during shift trials with square and ellipse irrelevant.

For NR shift the pair of cues which was irrelevant in training was made relevant on shift trials with one member consistently associated with reward. Two new cues replaced the previously relevant pair and these were made irrelevant. Thus, using the above example, circle and triangle which were irrelevant in training were now relevant for NR shift while yellow and green, made irrelevant, replaced red and blue.

Introduction of new cues was designed so that NR shift would not be impeded through partial reinforcement of the correct training response (House & Zeaman, 1962) and consequently to control for introduction of new cues in the R shift conditions.

Shift trials continued until *S* attained 9 correct of 10 successive responses or until a maximum of 50 trials was given. Upon completion of the shift series *S* was allowed to exchange the accumulated tokens for a

TABLE 2

Means and *SD*s of Transformed Trials to Criterion for Shifts

Relevant Dimension	Shift	Degree of Training		
		0 Trials	15 Trials	25 Trials
Color	Reversal	2.77 (1.05)	1.56 (1.23)	2.16 (0.97)
Form	Reversal	2.96 (1.07)	1.56 (0.69)	1.22 (0.56)
Color	Nonreversal	1.89 (1.38)	2.49 (1.87)	2.40 (1.31)
Form	Nonreversal	2.49 (1.16)	2.50 (0.99)	2.46 (0.97)

prize of his choice and he was told not to inform other *S*s about the game.

RESULTS

Trials to criterion, excluding the 10 criterion trials, was the response measure for training and shift performance and since these scores yielded skewed distributions a $\sqrt{x+.5}$ transformation was applied before analyses of variance were computed. Means and *SD*s of transformed scores for training are reported in Table 1 as a function of experimental conditions. The results of a three-way analysis of variance (Shift × Relevant Dimension × Degree of Training) provided no reliable effects, indicating preshift equivalance among groups.

For overtraining trials it was found that three *S*s made one, two, and four errors, respectively, while the other 61 *S*s who were overtrained made no errors during these trials. For shift performance no *S* required the maximum 50 trials. Transformed trials to criterion are reported in Table 2. A three-way analysis of variance yielded a reliable interaction of Shift with Degree of Training, $F\ (2,\ 84) = 4.65,\ p < .02$, while no other reliable source of variance was found. Since the interaction indicated a differential effect of overtraining on R and NR shift, *t* tests were computed between R and NR shift at each degree of training. After criterial training no shift difference was found ($t = 1.64$), while after either 15 or 25 overtraining trials R shift was found to require less trials than NR shift; $t = 2.16,\ p < .05$, and $t = 2.07,\ p < .05$, respectively.

A further analysis of R shift performance compared numbers of errors made in Trials 1–3 to those made in Trials 4–6 since it was previously reported (Furth & Youniss, 1964) that overtrained *S*s, but not *S*s trained only to criterion, showed a significant decline in errors from the first to the second 3 trials of the reversal series. In the present study there were 48 possible errors in a 3-trial block per degree of training condition. For *S*s given no overtraining a drop in errors from 36 to 27 was observed from the first to the second 3-trial block; this change proved to be nonreliable according to the Wilcoxon matched-pairs test. For *S*s overtrained 15 and 25 trials, changes in errors from 29 to 9 and from 30 to 10, respectively, indicated reliable ($p < .01$ in both cases) drops in errors from the first to the second 3-trial block.

DISCUSSION

The observed interaction of relative shift difficulty with degree of training obtained with trials to criterion agrees with the findings of Brookshire et al. (1961) and Mackintosh (1962) for rats as well as with Furth and Youniss' (1964) observation with children. Specifically, these data suggested a selective facilitation for R shift following overtraining, and this result was obtained in the present study along with the absence of a facilitating effect of overtraining on NR shift.

This interaction of type of shift with degree of training and comparisons on early reversal trials between overtrained and nonovertrained *S*s support a mediating-response model of discrimination learning. With training beyond criterion a dimension-specific response may have been strongly established so that during shift trials, once old responses were abandoned, new responses were made to cues within the initially relevant stimulus dimension. After criterial training this inferred dimension-specific mediator was probably less well established, hence new responses were not so readily directed to the relevant dimension. This interpretation can be considered compatible with either House and Zeaman's (1962) position or Kendler and Kendler's (1962) analysis of dimension-specific mediators. Since these authors have postulated that a mediating response which helps to identify the relevant dimension can be established through reinforcement of overt choices during training, it should follow that overtraining provides a mediator of greater strength than that established with criterion training only.

This interpretation, however, does not rule out other possible training effects on shifts. The immediate sharp decline in errors observed with overtrained *S*s in R shift suggests a rapid extinction of old overt choices after a change in reinforcement on early reversal trials. This change in reinforcement was probably more easily recognized as a cue for response change after a long series of 100% reinforced overtraining than after criterial training (Capaldi & Stevenson, 1957; Youniss & Furth, 1964a). Thus, two factors—a dimension-specific response coupled with observed change in reinforcement—might reasonably explain the alteration in reversal-shift difficulty as a function of overtraining.

Concerning extradimensional shift, it might be expected that with overtraining the dimension-specific response should have impeded performance since this mediator inappropriately identified irrelevant cues (cf. House & Zeaman, 1962). However, the cue for reinforcement change was available in this shift also, so that one effect could have counteracted the other, leaving extradimensional shift difficulty relatively unchanged across degrees of training.

Present data are in substantial agreement with a growing number of studies which demonstrated reversal facilitation after overtraining in children (Furth & Youniss, 1964; House & Zeaman, 1962; Youniss & Furth, 1964a, 1964b). This consistent overtraining effect of relative ease of extinguishing old responses and acquiring reversal responses supports a double-unit model of discrimination learning. Apparently overt choices on the preshift task are increasingly controlled by implicit responses established with increased training and not alone by individual S-R bond strengths, and subsequently the implicit responses mediate in shift performance.

REFERENCES

BROOKSHIRE, K. H., WARREN, J. M., & BALL, G. G. Reversal and transfer learning following overtraining in rat and chicken. *J. comp. physiol. Psychol.*, 1961, **54**, 98–102.

CAPALDI, E. J., & STEVENSON, H. W. Response reversal following different amounts of training. *J. comp. physiol. Psychol.*, 1957, **50**, 195–198.

FURTH, H. G., & YOUNISS, J. Effect of overtraining on three discrimination shifts in children. *J. comp. physiol. Psychol.*, 1964, **57**, 290–293.

HOUSE, B. J., & ZEAMAN, D. Reversal and nonreversal shifts in discrimination learning in retardates. *J. exp. Psychol.*, 1962, **63**, 444–451.

KELLEHER, R. T. Discrimination learning as a function of reversal and nonreversal shifts. *J. exp. Psychol.*, 1956, **51**, 379–384.

KENDLER, H. H., & KENDLER, T. S. Vertical and horizontal processes in problem solving. *Psychol. Rev.*, 1962, **69**, 1–16.

KENDLER, T. S., KENDLER, H. H., & WELLS, D. Reversal and nonreversal shifts in nursery school children. *J. comp. physiol. Psychol.*, 1960, **53**, 83–88.

MACKINTOSH, N. J. Effects of overtraining on a reversal and a nonreversal shift. *J. comp. physiol. Psychol.*, 1962, **55**, 555–559.

YOUNISS, J., & FURTH, H. G. Reversal learning in children as a function of overtraining and delayed transfer. *J. comp. physiol. Psychol.*, 1964, **57**, 155–157. (a)

YOUNISS, J., & FURTH, H. G. Reversal performance in children as a function of overtraining and response conditions. *J. exp. child Psychol.*, 1964, **1**, 182–188. (b)

(Received June 25, 1964)

another implies that prior information is available to the individual (assuming he does not use some external chance device such as flipping a coin). The information may be incorrect, a function of bias, set, even black magic, but some set of cues decides the individual. Once the decision has been made, numerous hypothetical mechanisms ranging in elegance from "rationalization" to "cognitive dissonance" can be invoked to account for a still greater increase in certainty prior to the first piece of information.

REFERENCES

ADAMS, S. K. A confidence scale defined in terms of expected percentages. *Amer. J. Psychol.*, 1957, **70**, 432–436.

ESTES, W. K. The problem of inference from curves based on group data. *Psychol. Rev.*, 1956, **53**, 134–140.

GRANT, D. Testing the null hypothesis and the strategy and tactics of investigating theoretical models. *Psychol. Rev.*, 1962, **69**, 54–61.

IRWIN, F. W., & SMITH, W. A. S. Further tests of theories of decision in an "expanded judgment" situation. *J. exp. Psychol.*, 1956, **52**, 345–348.

IRWIN, F. W., & SMITH, W. A. S. Value, cost, and information as determiners of decision. *J. exp. Psychol.*, 1957, **54**, 229–232.

IRWIN, F. W., SMITH, W. A. S., & MAYFIELD, J. F. Tests of two theories of decision in an "expanded judgment" situation. *J. exp. Psychol.*, 1956, **51**, 261–268.

(Received July 3, 1964)

Journal of Experimental Psychology
1965, 70, No. 4, 433–434

SUPPLEMENTARY REPORTS

FIXATION TIME AS A FUNCTION OF STIMULUS UNCERTAINTY

JAMES ALLISON
Indiana University

46 Ss were adapted to 1 of 3 levels of stimulus uncertainty, then tested at 1 of the 3 levels. Ss shifted upward fixated the test stimuli longer than Ss shifted downward, and this effect increased with the magnitude of the shift ($p < .01$).

Ordinarily, experiments showing that *Ss* select or orient toward stimuli which are relatively novel, complex, or uncertain (e.g., Berlyne, 1958; Fantz, 1958; Jones, 1964) do not manipulate systematically *S*'s pretest adaptation to the stimulus dimension. Frequently *S*'s adaptation level is uncontrolled, leaving little or no opportunity to determine for any *S* whether a test level represents an increase, a decrease, or no difference from his adaptation level. In the present experiment three sets of stimuli, ordered in uncertainty, were used as adaptation and test stimuli. Some *Ss* were adapted to Set 1 (low uncertainty), some to Set 2 (intermediate), and some to Set 3 (high). The dependent variable was the total time *S* fixated stimuli available during the test period. Those *Ss* shifted upward between adaptation and test were expected to fixate longer than *Ss* shifted downward, and this effect was expected to increase with the magnitude of the shift. That is, a monotone increase in total fixation time should occur across the following five groups, where the first number in each dyad identifies the adaptation set, the second the test set: (*a*) 3, 1; (*b*) 2, 1; 3, 2; (*c*) 1, 1; 2, 2; 3, 3; (*d*) 1, 2; 2, 3; (*e*) 1, 3.

Method.—The stimuli were several 1 × 4 in. bits of paper. Set 1 (low uncertainty) consisted of 42 identical white stimuli. Set 2 (intermediate) consisted of 14 whites, 14 identical blues, and 14 identical reds. Of the 42 stimuli in Set 3 (high) no two were identical, each forming a unique combination of Munsell hue, brightness, and saturation.

The stimuli were presented tachistoscopically, appearing successively in a 1 × 4 in. aperture against a homogeneous, dark gray ground. During the adaptation period each stimulus was illuminated for 5 sec., and the interstimulus interval was a 15-sec. period of darkness. During the test period *S* controlled the length of time each stimulus was illuminated by means of a telegraph key, in circuit with the lighting system and a .005-sec. Standard Electric clock.

During the adaptation period (14 min.) *S* was presented with one complete set of stimuli over 42 trials, one stimulus per trial. Following the adaptation period *S* rested for approximately 3 min. in a room adjacent to the experimental room. Upon returning for the test period, *S* was told that this time he would control the length of time each stimulus was illuminated, by pressing the telegraph key. When *S* was through looking at a given stimulus he would release the key, blacking out the tachistoscope. The *E* would then move the next stimulus into position, signal to *S*, who would press the key again, etc. The instructions emphasized that *S* was free to look at any stimulus for as long as he pleased, and that he was free to end the experiment whenever he liked. The *E* recorded the length of time *S* illuminated each stimulus, and summed these times to obtain total fixation time.

Stimuli in Set 3 were presented in a random order. Set 2 was randomized within blocks of 15, 15, and 12 trials, with the restriction that the three colors appear equally often within each block. One order of presentation was used for all *Ss*, and for both periods.

The *Ss* were 46 students from introductory psychology courses at the University of Michigan. The experimental design was a

TABLE 1

MEDIAN TOTAL FIXATION TIME (SECONDS)

Adaptation Set	Test Set			
	1	2	3	
1	46	40	72	50
2	14	34	48	36
3	22	34	37	30
	32	34	55	

3 × 3 factorial with three levels of stimulus uncertainty during adaptation, and three during the test. The number of *S*s in each cell varied from four to six.

Results and discussion.—The *S*s shifted from a lower to a higher level of uncertainty fixated longer than those shifted from a higher to a lower level, and this effect increased with the magnitude of the shift. Medians for the five appropriate groups were 22, 25, 37, 47, and 72 sec. (Kendall's tau = .25, $z = 2.47$, $p < .01$). The fundamental point of this paper is illustrated by the marginal entries in Table 1, which reflect the main effects of the adaptation stimuli and the test stimuli. It is clear that each of these effects is partially confounded with the shift effect. For example, Test Set 3 represents an average upward shift by one level, while Test Set 1 represents an average downward shift by one level. Because the confounding is not complete, neither main effect takes maximum advantage of the shift information.

REFERENCES

BERLYNE, D. E. The influence of complexity and novelty in visual figures on orienting responses. *J. exp. Psychol.*, 1958, **55**, 289–296.

FANTZ, R. L. Pattern vision in young infants. *Psychol. Rec.*, 1958, **8**, 43–48.

JONES, A. Drive and incentive variables associated with the statistical properties of sequences of stimuli. *J. exp. Psychol.*, 1964, **67**, 423–431.

(Received June 1, 1964)

Journal of Experimental Psychology
1965, 70, No. 4, 434–436

FURTHER STUDY OF EFFECTS OF LOW DRIVE STATES ON COMPETING RESPONSES[1]

CHARLES Y. NAKAMURA AND WILLIAM E. BROEN, JR.

University of California, Los Angeles

The effect of drive reduction, within the range of low drives, on the probability of a competing response was further explored. Ss were college males. Prior evidence that drive reduction facilitates a competing response was supported, but only under conditions where response strengths met certain theoretical boundary requirements. The results indicate that it is critically important in this type of experimentation to provide considerable control for establishing the strengths of the relevant responses.

The hypothesis that a reduction of drive below an already low drive level would lead to an increase in the probability of a competing response was supported by a previous experiment (Nakamura & Broen, 1965). The present experiment was intended to test the stability of the findings. Certain procedural modifications were introduced to increase the efficiency of experimentation.

In the earlier experiment, *S*s were initially trained with reinforcement probabilities of .5 for the competing response and of .8 for the dominant response with the objective of establishing a clearly dominant response, which was not so superior in habit strength that the probability of an increment in the competing response would be negligible under the amount of drive reduction used. If the dominant response were too strong, the prediction of an increase in probability of the competing response may not apply for at least two reasons: (*a*) The strengths of the dominant and competing responses may be so far apart that the relative response probabilities would not be altered by small changes in drive level; and (*b*) If the competing response is initially very weak, it could go below threshold with reduction in drive level and the response probability could thus decrease, or not change at all, instead of being facilitated by the drive reduction. In order to assure an adequate dominant response, a learning criterion of 70% dominant responses by the end of the training trials was selected. These procedures, used in the previous experiment, resulted in a high rate of rejection of *S*s. Therefore, the present experiment used simpler responses so that more *S*s might reach criterion.

Method.—The *S*s were 52 males from an introductory psychology course who reached the criterion of 70% dominant responses over the last 30 trials of training. Not used were 18 *S*s who did not reach criterion.

---

[1] The authors wish to thank Fred Shima for running *S*s and tabulating data. This research was supported by funds from the University Research Committee.

Except for the modification of the finger-response unit, the apparatus was the same as in the previous experiment (Nakamura & Broen, 1965). The response unit comprised a plate with two $\frac{3}{8}$-in. diameter holes. Each hole contained a contact pin that activated a transistor relay circuit which operated impulse counters.

Throughout the experiment *S* responded while lying on a couch. Starting with forefinger on a marker point between the two holes, *S* made a finger movement, right or left, to contact either pin. These alternative responses were presumed to be conceptually simpler than the move or not-move responses, in the previous study, which could begin from either the right or left pin on each trial.

The *S* was instructed to make the correct response following each start signal (high-frequency tone). Each correct response would be followed by another tone (low frequency) and incorrect responses by no tone. The *S* received 150 training and 120 test trials. Training and test sessions were separated by 4 min. during which the relaxation group received relaxation exercises and the control group solved a mental task. The reinforcement conditions remained the same over both training and test trials. All *S*s were told that the task in the second session was to be identical to that in the first. Finally, each *S* was asked to rate his degree of relaxation in each of the preceding two sessions on a 9-point rating scale ranging from not relaxed to very relaxed. A detailed account of these procedures is given elsewhere (Nakamura & Broen, 1965).

Results and discussion.—The Relaxation ratings provide evidence that the treatment was effective. The *S*s in the Relax groups rated themselves as more relaxed in the second compared to the first session than did the Nonrelax *S*s, *F* (1, 48) = 11.65. This finding is consistent with the trend of the results in the previous study (Nakamura & Broen, 1965).

The use of simpler responses did increase the efficiency of training. The criterion was met by 74% of the *S*s compared to 36% in the earlier study. However, this also produced the more serious problem discussed above of too strong a dominant response, or conversely, a very weak competing response. Under the theoretical consideration, it was apparent that the competing response may be too weak to permit an adequate test of the hypothesis. Accordingly, it was decided to divide *S*s into two levels of competing response strengths; very weak (0–10%) and moderate (11–30%). The range of 0–10% as weak was selected prior to examination of the test-trials data. Moreover, since the analysis of Response Type was not significant, it was excluded from the subsequent analyses to simplify the division of *S*s into the two levels of competing response strengths, based on the last 30 trials of training. This resulted in 9 *S*s in the Weak level and 17 *S*s in the Moderate level of each treatment.

The performances of the Treatment groups at the two levels of training are shown in Fig. 1. The mean percentages of competing response on training trials are at the left side of the graph; four blocks of 30 test trials are presented by the connected points. As can be seen by the curves, there is little indication of an overall Treatment effect and this was verified by several analyses. Four analyses-of-variance tests using difference scores (*a*) between the mean of the last 30 training trials and the mean of the 120 test trials, (*b*) between training trials and Block 1 test trials, (*c*) between training trials and Block 4 test trials, and (*d*) between Block 1 and Block 4 test trials yielded no significant *F*s for the Treatment main effect. However, the Treatments × Levels interactions warranted further attention. The T × L interaction was significant, *F* (1, 48) = 10.34, for Analysis *c* above. Examination of the curves in Fig. 1 suggests that the source of this interaction is within the Moderate level of response strength where the boundary conditions of the theoretical account are presumably best satisfied, and thus where the prediction should receive strongest support. Within this Moderate level the Relax group ends the test trials with mean competing response higher than the Nonrelax group after a gradual increase in the

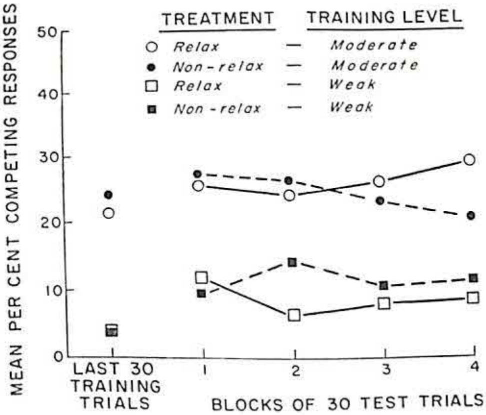


FIG. 1. Mean percentage of competing responses in the last 30 training trials and over blocks of test trials for groups of *S*s who attained weak and moderate strengths of the competing response by the end of training.

competing response, while the latter group shows a corresponding decrease. This locus of the above effect was confirmed by the significant result, $F(1, 32) = 11.50$, of a specific comparison of these two Treatment groups within the Moderate level, while a similar test within the Weak level was not significant. The T × L interaction in Analysis *d* above between the means of the Block 1 and Block 4 test trials was just short of the .05 level of significance, $F(1, 48) = 3.99$. However, in the specific comparison of the Relax and Nonrelax groups within the Moderate level, the result was again significant, $F(1, 32) = 4.45$. In contrast the comparison within the Weak level was not significant.

Together, the results of these tests indicate that there was a Treatment effect in the predicted direction of increase in competing responses as a function of lowered drive in *S*s that had attained a moderate strength of the competing response by the end of training, but this effect did not appear in *S*s in whom there was established a weak strength of the competing response. Thus, these findings are consistent with the previous research and are compatible with the theoretical expectations that where the dominant response is very strong and the competing response relatively weak, the latter may not be facilitated by a reduction in drive level because of too large a difference between the response strengths or because the strength of the competing response may go below threshold.

On the basis of a subsequent experiment, a note of caution may be in order on the necessity for precise manipulation of drive levels in this type of experimentation. The experiment, designed to permit larger magnitudes of drive reduction than in the present study, yielded nonsignificant results. Scores on the Relaxation Rating scale gave evidence that the relaxation treatment was not effective under the conditions used (bright room lighting) to increase the training drive level. Unfortunately, the findings deserve no further discussion here since the results were confounded by procedural changes in the finger responses.

REFERENCE

Nakamura, C. Y., & Broen, W. E., Jr. Facilitation of competing responses as a function of "subnormal" drive conditions. *J. exp. Psychol.*, 1965, **69**, 180–185.

(Received June 6, 1964)

Journal of Experimental Psychology
1965, 70, No. 4, 436–437

RUNNING SPEED IN THE RAT AS A FUNCTION OF SHOCK LEVEL AND COMPETING RESPONSES[1]

GEORGE A. CICALA AND J. R. COREY
University of Delaware

3 groups of rats were trained to escape shock in a straight alley under 3 drive conditions: 90 v. (N = 7), 60 v. (N = 7), and alternated 60 and 90 v. (N = 9). Under either the alternated or nonalternated conditions, removal of competing-response trials did not serve to reduce running-speed differences as a function of shock level. The results were related to those of Cotton (1953) and Cicala (1961) and were discussed in relation to Estes' (1958) drive-stimulus theory.

Recent studies (Cicala, 1961; Cotton, 1953) have demonstrated that the removal of competing-response trials may reduce differences in asymptotic runway performance as a function of hunger-drive level. Estes (1958) has interpreted such results as favoring his drive-stimulus theory, in that removal of competing responses reduces the differential effects of extraneous stimuli present under different deprivation conditions.

Cicala (1961) has demonstrated that this effect is contingent upon training *S*s under alternated drive conditions, since groups trained under separate, nonalternated conditions continued to exhibit differences in performance as a function of drive level, even when competing responses were not considered in the analysis. The purpose of the present study was to assess the generality of Estes' drive theory for an escape-responding situation, utilizing alternated and nonalternated drive levels and observing the effects of competing responses.

Method.—A 6-ft., grid-floored, Plexiglas and aluminum straight runway with interchangeable goal boxes was used. Shock was delivered to the grids of the start box and

---

[1] This research was supported by Grant MH-06725-02 from the Public Health Service and represents, in part, the junior author's MA thesis submitted to the faculty of the University of Delaware.

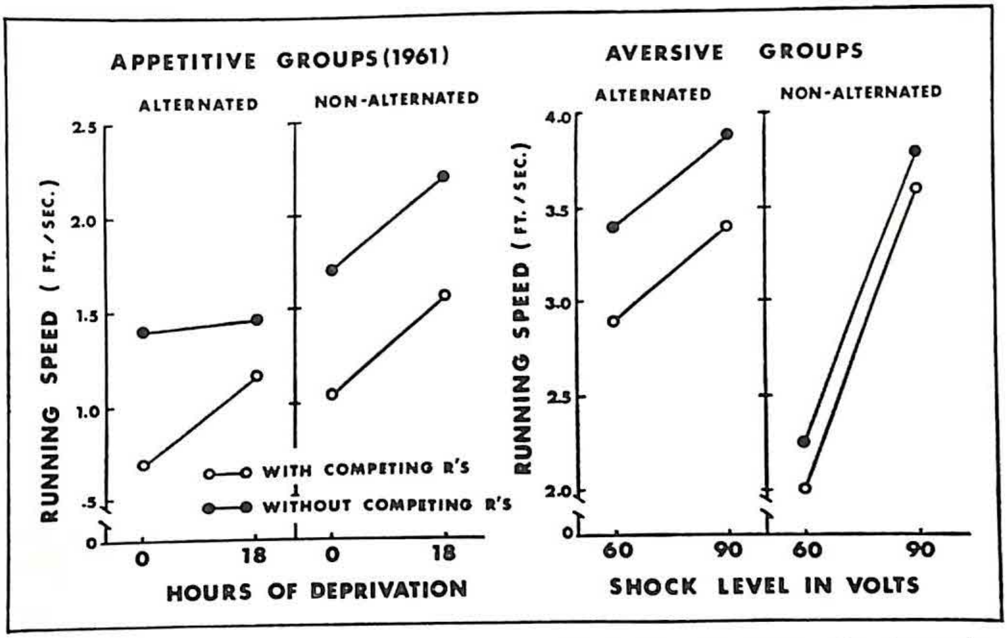


FIG. 1. Comparison between Cicala's (1961) data and the results of the present study. (Running speeds as a function of drive level and the presence of competing-response trials are plotted in each case.)

main runway by means of a matched-impedance, scrambled ac shock source. A trial was initiated by lifting the guillotine door of the start box and presenting the shock. While *S*s' running time was recorded by means of a photocell-timer circuit, one or more *E*s noted the presence of competing responses (cf. Cotton, 1953). The two *E*s who observed on 20% of the trials reached 95% agreement as to the presence of competing responses.

The *S*s were 23 male Wistar albino rats, 95–120 days old. Fourteen *S*s were trained under the nonalternated condition; 7 at 60 v., 7 at 90 v. Nine *S*s were presented 60 and 90 v., alternated from trial to trial. All received 10 trials a day with a 45-sec. intertrial interval for 4 consecutive days.

Results and discussion.—Running times were converted into speed scores. Analyses of the acquisition curves revealed that all groups reached asymptotic performance by the thirtieth trial. Further analysis of asymptotic speeds (Alternated Trials 33–40, Nonalternated Trials 37–40) was undertaken with and without the inclusion of competing response trials. Differences in running speed as a function of shock level with or without competing responses were significant at less than the .01 level in every case.[2]

A comparison of the results of the present study with those of Cicala (1961) appears in Fig. 1. Whereas the removal of competing-response trials serves to eliminate the effects of an alternated food-deprivation level upon performance in the 1961 study, no such effect occurs in either condition of the present study. It is suggested that rapidly alternated shock levels may afford better comparison for drive-discrimination learning than hunger-drive levels, since the latter must be alternated more slowly.

The results of the present study do not support Estes' statistical theory as presently formulated, but are consistent with Hull's two-factor theory of drive. Estes might explain these results on the basis of different amounts of stimulus change at different drive levels, an explanation which would derive some support from the work of Campbell and Kraeling (1953). However, the effects of incentive magnitude are yet to be demonstrated in the aversive situation when competing responses are observed.

REFERENCES

CAMPBELL, B. A., & KRAELING, D. Response strength as a function of drive level and amount of drive reduction. *J. exp. Psychol.*, 1953, **45**, 97–101.

CICALA, G. A. Running speed in rats as a function of drive level and presence or absence of competing response trials. *J. exp. Psychol.*, 1961, **62**, 329–334.

COTTON, J. W. Running time as a function of amount of food deprivation. *J. exp. Psychol.*, 1953, **46**, 341–357.

ESTES, W. K. Stimulus-response theory of drive. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1958.* Lincoln: Univer. Nebraska Press, 1958. Pp. 35–69.

(Received June 4, 1964)

---

[2] Nonalternated groups, Mann-Whitney *U* (all trials, $N_{60} = 28$, $N_{90} = 28$) = 106, *U* (competing-response trials eliminated, $N_{60} = 18$, $N_{90} = 24$) = 42.5; alternated group, Wilcoxon grouped replicates *T* (all trials, $N = 72$) = 117, *T* (competing-response trials eliminated, $N = 48$) = 66.

Journal of Experimental Psychology
1965, 70, No. 4, 438–440

EFFECTS OF DELAY AND MODE OF PRESENTATION OF EXTRA CUES ON PURSUIT-ROTOR PERFORMANCE[1]

LAWRENCE KARLIN

New York University

The effects on pursuit-rotor performance of using an extra cue and varying its mode of presentation were investigated in order to examine the reasons for previous results that were in conflict. An extra cue was found to be effective but variations in the interval of delay were not significant.

In a pursuit-rotor task, Reynolds and Adams (1953), using auditory clicks which were sounded at equally spaced intervals for each corresponding interval of time that *S* remained continuously on target, found that the click interval made a difference—a $\frac{1}{2}$-sec. interval yielding consistently better performance during training than a number of shorter and longer intervals. Archer and Namikas (1958), however, found that variations in delay were not significant but the experimental conditions of their study were different in certain respects. One difference was the use of a continuous tone as compared to a discrete click. It was felt that the reinforcing properties of tone and click might be different. Archer and Namikas also used considerably fewer training trials. Finally, the pursuit-rotor used by Archer and Namikas turned on a 5-in. radius unlike that of Reynolds and Adams which turned on a radius of about 3.19 in.

The present study compared the effects of discrete and continuous modes of stimulus presentation on the delay variable and used a pursuit-rotor and number of training trials more comparable to that of Reynolds and Adams.

Method.—Eighty right-handed male college students divided equally among the five conditions served as *S*s.

A Lafayette pursuit rotor, Model 203, was used with a .75-in. target turning on an orbit of 3.25 in. It was tracked at 60 rpm using a hinged stylus with a flat tip of $\frac{3}{16}$ in. diameter. Time on target was measured by a Hunter klockounter. Auditory stimuli were 1,000-cycle tones of moderate intensity presented to *S* through headphones. A special circuit actuated when *S* was on target, presented this tone continuously after a preset delay, or in .025-sec. bursts at regular intervals equal to the initial delay. The turntable speed was calibrated with a stroboscope prior to and after each session.

All conditions were administered equally often at different hours of the day and equally often throughout the experiment. The control group wore headphones but heard no tone. The four experimental groups were as follows: .5 and .2 sec. continuous, and .5 and .2 sec. discrete. The continuous tone was sounded following either a .2- or .5-sec. interval of continuous time on target and kept sounding as long as *S* remained on target. The discrete tone sounded when *S* maintained continuous contact for the requisite amount of time and was repeated for equal intervals of continuous time on target thereafter.

The *S*s were encouraged to do their best. All groups received comparable instructions except that no mention was made of extra cues to the control group. There were 52-training and 15-transfer (no extra cue) trials administered to all groups. Trials were 30 sec. long with 15-sec. intertrial intervals. Following Trial 52 there was a 1-min. rest period. Experimental *S*s were then told that the extra cue would be removed.

Results.—Figure 1 shows that all groups were approximately equal at the beginning of training, but the experimental groups were superior to the control group by the end of training and during the transfer trials. Three of the groups yielded trends with considerable overlap although the .2 continuous group was superior to the others during training and transfer.

Trend analyses (Edwards, 1960) of the training curves comparing each experimental group to the control group yielded significant results for the linear component in all four comparisons at better than the .001 level. None of the experimental trends was significantly different from any other. The means of the last five training trials and of the first and last five transfer trials for each group are shown in Table 1. Variance analyses of

---

[1] The writer wishes to thank Richard Kestenbaum who collected the data and helped in its analysis.

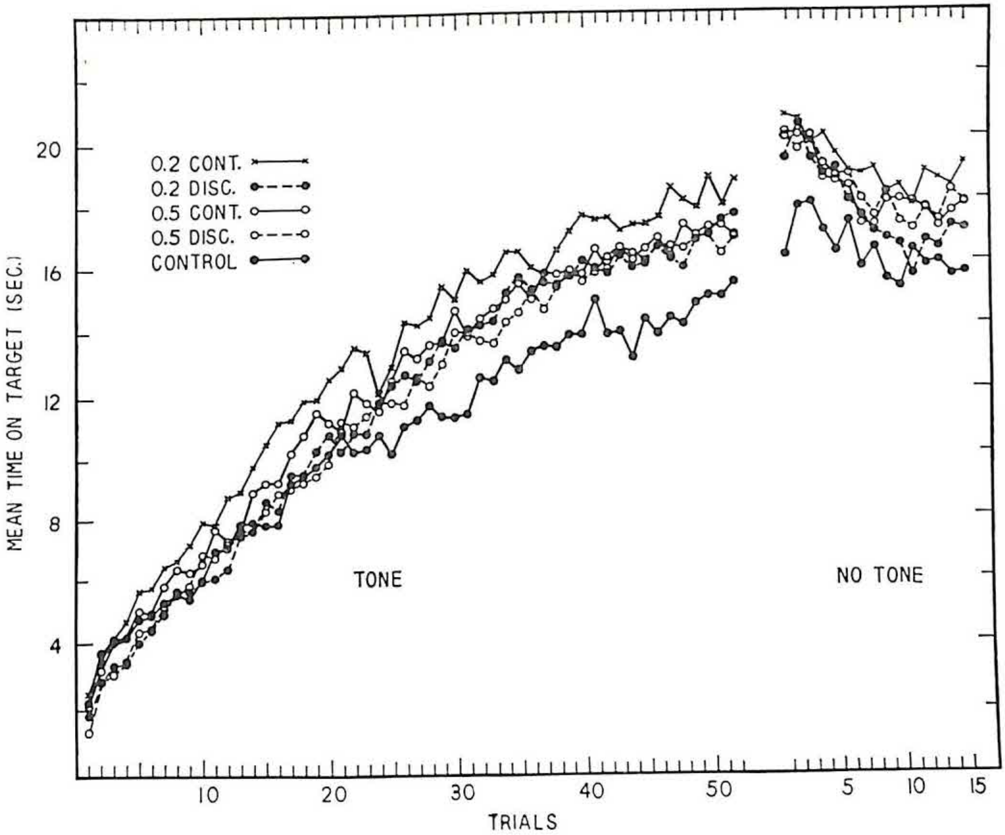


FIG. 1. Mean time on target by trial for each group. (Each point represents the mean of 16 *Ss*.)

each of these three sets of means for the four experimental groups and for all five groups yielded highly insignificant results. A Duncan range test (Edwards, 1960) applied to the training means yielded significant results (.05 level) only for the difference between the means of the .2 continuous and control conditions. In the first five transfer trials, all experimental means differed significantly from the control mean (.05 level) but not from each other. The last five transfer trials again yielded significant results only for the difference between the means of the .2 continuous and control conditions.

Discussion.—In the present study, .2- and .5-sec. delay intervals were used to maximize the difference in effectiveness of the extra cue since Reynolds and Adams (1953) using intervals ranging from .1 to 2.0 sec. found that a .5-sec. interval yielded the best results while the .2-sec. interval was about the poorest, although it was better than no cue at all. The .2 and .5 conditions of the present study do not support the results obtained by Reynolds and Adams since neither acquisition nor transfer curves suggest that the .5 interval (discrete or continuous) is superior. Hence there is some support in the present results for the conclusion reached by Archer and Namikas (1958) that "delay of information feedback is not an effective variable." However, it is worth noting that there are certain differences between the present study and that of Reynolds and Adams (1953). First, even though an effort was made to match pursuit rotors, the present *Ss* score on the

TABLE 1

MEAN SCORE IN SECONDS FOR SELECTED FIVE-TRIAL BLOCKS DURING TRAINING AND TRANSFER FOR ALL GROUPS

Group	Trial		
	48–52	53–57	63–67
.2 Continuous	18.36	20.10	18.59
.2 Discrete	16.95	19.49	16.49
.5 Continuous	17.05	19.56	17.79
.4 Discrete	16.61	19.54	17.63
Control	15.00	17.00	15.84

average about 2 sec. or 7% less. Possibly this is due to some differences (motivation, etc.) between *S*s in the two studies rather than in the task. Secondly, we did not use the matching technique of Reynolds and Adams (1953) who used five preliminary trials for this purpose. Since our time per *S* was limited, we preferred to use more transfer trials. Reynolds and Adams using only five transfer trials found that the delay variable was not significant.

Perhaps further replication along lines suggested by the above differences would support Reynolds and Adams but the present results suggest that this is not likely. In fact it is the .2-sec. continuous condition which is consistently superior throughout training and transfer. Since Archer and Namikas (1958) also report this tendency, it may be worth further study.

REFERENCES

Archer, E. J., & Namikas, G. A. Pursuit rotor performance as a function of delay of information feedback. *J. exp. Psychol.*, 1958, **56**, 325–327.

Edwards, A. L. *Experimental design in psychological research.* (Rev. ed.) New York: Rinehart, 1960.

Reynolds, B., & Adams, J. A. Motor performance as a function of clock reinforcement. *J. exp. Psychol.*, 1953, **45**, 315–320.

(Received June 15, 1964)

Journal of Experimental Psychology
1965, 70, No. 4, 441–442

Beginning with the first issue of Volume 69 and continuing through the remainder of that volume and Volume 70 (1965), the titles and authors of accepted papers will be listed here following Supplementary Reports. It is being supported on an experimental basis by the APA Project on Scientific Information Exchange in Psychology, and at the end of the year, the outcome of this trial will be evaluated and consideration given the advisability of continuing the listing.

This listing plus those published in the preceding 1965 issues are a portion of the entire backlog of manuscripts accepted by this journal. Such listing will allow readers to become aware of research many months in advance of journal publication.

Manuscripts Accepted for Publication in the *Journal of Experimental Psychology*

Probability Matching in Concept Identification: J. Brown Grier* and Robert Bornstein: Department of Psychology and Social Anthropology, University of Louisville, Louisville, Kentucky 40208.

Transfer Effects and Response Strategies in Pattern versus Component Discrimination Learning: Morton P. Friedman*: Department of Psychology, University of California, Los Angeles, California 90024.

Response Decrement in Extinction and Counterconditioning as a Function of the Number of Reinforced Guessing Responses: Edward L. Trembley and John E. Nygaard*: Department of Psychology, California State College, Long Beach, California 90804.

Information Encoding and Decision Time as Variables in Human Choice Behavior: Louis M. Herman* and Harry P. Bahrick: Department of Psychology, Queens College, Flushing, New York 11367.

Visual Detection of Compound Motion: L. T. Alexander* and A. S. Cooperband: System Development Corporation, 2500 Colorado Avenue, Santa Monica, California 90406.

Continuing Memory and Information Processing: A. F. Sanders* and J. W. van Borselen: Department of Psychology, University of Michigan, Ann Arbor, Michigan 48104.

Multivariate Analysis of Preference for Variability: Harry Munsinger*: Department of Psychology, University of Illinois, Urbana, Illinois 61803.

Two Replications of Avoidance Conditioning of the GSR: H. D. Kimmel,* H. S. Sternthal, and H. Strub: Department of Psychology, University of Florida, Gainesville, Florida 32603.

Intensity-Time Relationship and Perceived Shape: H. W. Leibowitz,* Sharon E. Toffey, and John L. Searle: Department of Psychology, Pennsylvania State University, 117 Burrowes Building, University Park, Pennsylvania 16802.

Role of Stimulus Labeling in Stimulus Generalization: David R. Thomas* and Alberta DeCapito: Department of Psychology, Kent State University, Kent, Ohio 44240.

Transfer as a Function of Time to Mediate: Norman E. Spear,* Peter J. Mikulka, and Marvin Podd: Department of Psychology, Rutgers State University, New Brunswick, New Jersey 08903.

Serial List Items as Stimuli in Paired-Associate Learning: Sheldon M. Ebenholtz*: Department of Psychology, Connecticut College, New London, Connecticut.

Effect of Interpolated Extinction and Level of Training on the "Depression Effect": John R. Vogel, Peter J. Mikulka, and Norman E. Spear*: Department of Psychology, Rutgers State University, New Brunswick, New Jersey 08903.

All-or-None Learning of Attributes: Albert S. Bregman* and David W. Chambers: Center for Cognitive Studies, Harvard University, 33 Kirkland Street, Cambridge 38, Massachusetts.

Effect of Intrastimulus Agreement and Conflict Upon Conditioned Discrimination and Reversal of the Eyelid Response: C. Michael Levy*: Department of Psychology, University of Florida, Gainesville, Florida 32603.

Stimulus Intervals, Stimulus Durations, and Difficulty Level in Paired-Associates Learning: Calvin F. Nodine* and Barbara F. Nodine: Department of Psychology, Carnegie Institute of Technology, Pittsburgh, Pennsylvania 15213.

---

* Asterisk indicates author for whom address is supplied.

Analysis of Learning Rate and Sampling Probabilities in a Contingent Reinforcement Situation: John G. Borkowski*: Department of Neurology, University Hospitals, University of Iowa, Iowa City, Iowa.

Task Predictability and the Development of Tracking Skill under Extended Practice: Merrill Noble,* Don Trumbo, Lynn Ulrich, and Kenneth Cross: Department of Psychology, Anderson Hall, Kansas State University, Manhattan, Kansas 66504.

Input and Output Interference in Short-Term Associative Memory: Endel Tulving* and Tannis Y. Arbuckle: Institute of Human Learning, University of California, 2241 College Avenue, Berkeley, California 94720.

Verbal Repetition, Set, and Decision Latency: William E. Gumenik* and Edward S. Perlmutter: Department of Psychology, University of Toledo, Toledo 6, Ohio.

5 *OUTSTANDING TEXTS* *from* BUTTERWORTHS

MENTAL MECHANISMS

by Henry P. Laughlin, M.D.

The most thorough presentation of clinical and theoretical data on the six major mental mechanisms currently available in any reference. An extremely popular text, widely used for student instruction, as well as for patient guidance during analysis.

CONTENTS: Repression. Sublimation. Identification. Inversion. Compensation. Restitution. Index.

272 pages. $7.50. 1963.

PSYCHOGENIC PSYCHOSES

by Paul M. Faergeman, M.D.

A valuable treatise on an important aspect of psychiatric research about which little has previously been written in English, i.e., the practical problem of diagnosis in the borderlands between the neuroses and the psychoses, and between schizophrenia and other mental disorders. An authoritative analysis which will assist every psychiatrist and clinical psychologist in his daily activities.

268 pages. $8.50. 1963.

The one text which must be read by every psychiatrist, psychologist, and student of mental disorders . . .

THE NEUROSES

by Henry P. Laughlin, M.D.

The publication of THE NEUROSES marks the beginning of a period when formal psychiatric study no longer need be classified as "heavy reading." Dr. Laughlin's presentation is specifically directed to the requirements of the Board Candidate in Psychiatry, for whom this text is essential reading during study and before examination.

The unusual ease of readability evident throughout this book is distinctly unique. For everyone interested in emotions, motivations and behavior, THE NEUROSES is unquestionably the best such text available and should be considered essential reading.

Approx. 1000 pages. Approx. $20.00. Due late 1965.

ORDER NOW FOR IMMEDIATE DELIVERY ON PUBLICATION—SOON TO BE PUBLISHED!

DETENTION OF THE MENTALLY DISORDERED

by Barry B. Swadron, LL.B., LL.M.

In this book the author takes the reader through the labyrinth of Canadian legislation on the mentally disordered as it varies from province to province. American psychiatrists will obtain considerable benefit from Mr. Swadron's reasoned analysis, criticism and rationalization of the current Canadian situation, which offers so many parallels to similar problems in the United States.

502 pages. $15.00. 1964.

CRIMINAL RESPONSIBILITY AND MENTAL ILLNESS

by F. A. Whitlock, M.D., M.R.C.P., D.P.M.

An important text dealing with the problem of presenting such matters as psychopathy, alcoholism, and disturbances of consciousness to the purist or layman. The legal and official government references are of purely British application, but the medical material is of universal interest, and will be easily understood by the interested layman.

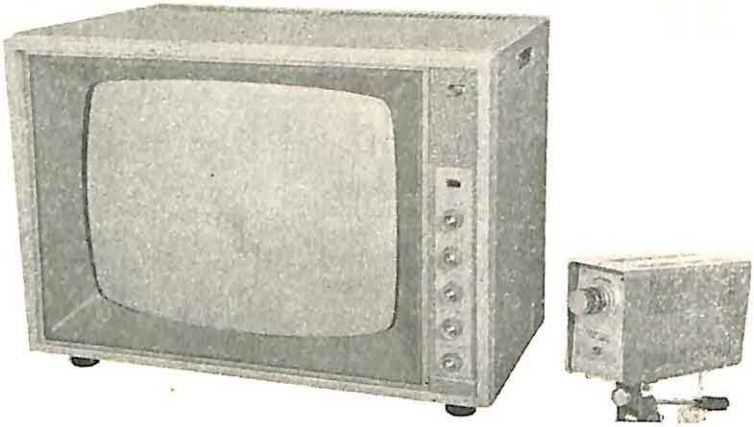
156 pages. $8.00. 1963.

ORDER THROUGH YOUR BOOKSELLER OR DIRECTLY FROM:

BUTTERWORTH INC.

7300 PEARL STREET WASHINGTON, D.C. 20014

SCIENTIFIC PROTOTYPE



LOW COST CLOSED-CIRCUIT TV

FOR CLINICAL AND RESEARCH APPLICATIONS

The quality and extremely low cost of this new closed circuit TV camera opens many new avenues of research to the clinician and researcher.

The camera offers a number of unique features. Two outputs are standard. One is the usual 4mc output to a closed circuit moniter. The other is a Channel 2 thru 6 output that can be connected to any commercial TV receiver. This enables the investigator to avoid the necessity of buying a special receiver for use with the camera.)

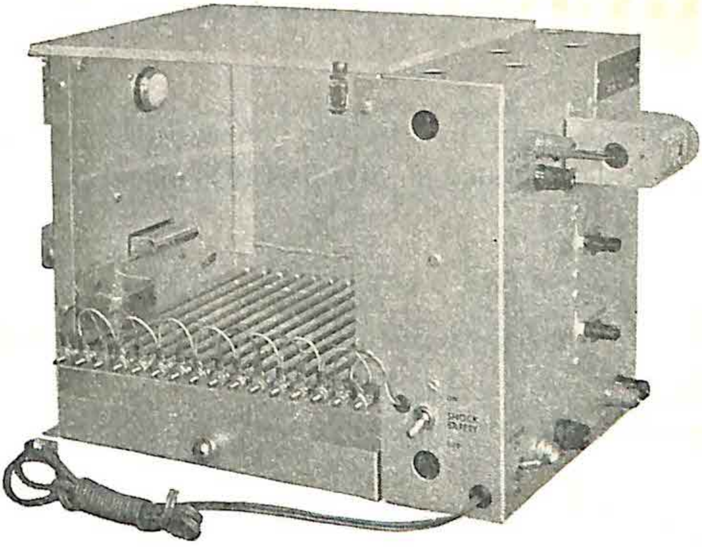
The camera is supplied with a standard f1.9., 25 mm C mount lens that will operate under 30 ft. candle ambient light conditions (less than normal room illumination).

A full selection of accessories is available including wide angle, manual and automatic zoom and telephoto lenses, tripod adapters, cables, line amplifiers, and transformers, pan and tilt equipment, etc.

Delivery is off-the-shelf. Price, only $450.00.

SCIENTIFIC PROTOTYPE MFG. CORP.
615 W. 131 Street, New York, N. Y. 10027
Telephone: (212) 368-6000

SCIENTIFIC PROTOTYPE



MODEL A-105 STUDENT LAB OR DEMONSTRATION BOX ANIMAL CAGE WITH MANY FEATURES BY SCIENTIFIC PROTOTYPE

This Scientific Prototype unit combines animal cage, pellet feeder, power supply, light and shock source intensity controls, and counter. The unit is completely wired, mounted on a base, and ready to operate.

The rat box is approximately 9 × 8⅝ × 8. Grids are stainless steel. Top and sides are ¼" clear Plexiglass. The stainless steel pedal lever is counter-balanced to permit sensitivity adjustments. Dropping tray is anodized aluminum.

Of special interest is the inclusion of self-contained cue light and shock source intensity controls.

Price . . . $259.00 F.O.B. N.Y.C. . . . quantity discounts. Full information available.

SCIENTIFIC PROTOTYPE MFG. CORP.

615 W. 131 Street, New York, N. Y. 10027
Telephone: (212) 368-6000

MICROFORM

All journals published by the American Psychological Association are being made available on MICROFILM or MICROCARD.

Journal of Consulting Psychology

American Psychologist **Psychological Monographs**

Journal of Applied Psychology

Journal of Personality and Social Psychology

Journal of Educational Psychology **Psychological Review**

Psychological Bulletin **Journal of Experimental Psychology**

Journal of Comparative and Physiological Psychology

Comtemporary Psychology

Journal of Abnormal Psychology **Psychological Abstracts**

Available only in volume units; no single issues.

For MICROFILM, order from:
UNIVERSITY MICROFILMS, INC.
313 North First Street
Ann Arbor
Michigan

For MICROCARD, order from:
J. S. CANNER & COMPANY, INC.
Microcard Division
618 Parker Street
Roxbury 20, Massachusetts

Journal of
Experimental Psychology

Vol. 70, No. 5 November 1965

SIGNAL-DETECTION THEORY AND SHORT-TERM MEMORY[1]

BENNET B. MURDOCK, Jr.[2]

University of Vermont

This study attempted to apply signal-detection theory to short-term memory by testing the high-threshold concept of associative strength. On each trial a list of 6 A-B pairs was presented once, then 1 of these 6 pairs was tested for recognition. On the recognition test either A-B (a proper pair) or A-X (an improper pair) was presented; S had to make a binary (yes-no) decision plus a confidence rating. From these data ROC curves were plotted, and they resembled the curvilinear functions of signal-detection theory more than the linear function required by the high-threshold concept. These results call into question the use of a high-threshold concept to explain findings from studies of 1-trial learning.

Recent developments in signal-detection theory have provided a "new look" at many problems in perception (Swets, Tanner, & Birdsall, 1961) and attention (Broadbent & Gregory, 1963). Basically, signal-detection theory makes possible a separation of the criterion and the sensitivity of *O* and provides some powerful tools for the analysis of these processes. It seemed likely that these methods could profitably be applied to problems in short-term memory, and the present study is such an attempt.

Specifically, the present study is an attempt to test the "high-threshold" concept in short-term memory. Underwood and Keppel (1962), in an article on one-trial learning, have suggested a threshold of associative strength below which no response will occur. Thus, results from studies which use an RTT paradigm can be explained; items above the threshold on T_1 will tend to be above the threshold on T_2 while items below the threshold on T_1 will tend to be below the threshold on T_2. However, as Estes (1962) has pointed out, recent threshold studies do not provide much support for this strong concept. The purpose of the present experiment was to try to make a direct test of the high-threshold concept in short-term memory itself.

---

[1] This investigation was supported by Public Health Service Research Grant MH 03,330 from the National Institute of Mental Health, Public Health Service. The writer is endebted to Donald A. Norman for helpful comments and suggestions. Cynthia M. Fischer assisted in collecting and analyzing the data.

[2] Now at the University of Toronto.

The procedure used in the present experiment followed directly along the lines of signal-detection theory. Short-term memory was tested by a recognition test supplemented with confidence ratings. As in previous studies (e.g., Murdock, 1963) on each trial a list of six paired associates was presented, then retention of one of the six pairs was tested. Here, however, retention was tested by recognition; *S* was shown either A-B (i.e., one of the pairs from the list) or A-X (i.e., not one of the pairs from the list) and asked to give both a yes-no judgment and a confidence rating. As has been pointed out (e.g., Egan, Schulman, & Greenberg, 1959; Pollack & Decker, 1958), with confidence ratings one has the necessary information to plot ROC curves. The question of interest, then, is whether the resulting ROC curves are the linear function demanded by the high-threshold concept (Swets, Tanner, & Birdsall, 1961).

PROCEDURE

All *S*s were tested individually and were given 54 trials on each of four sessions. Each trial consisted of a presentation of six A-B pairs followed by a recognition test for one of the six pairs. The recognition test consisted of presenting either an A-B or an A-X pair; *S* had to respond either "yes" or "no" and, in addition, give a confidence judgment. An A-B pair would be a proper pair while an A-X pair would be an improper pair; that is, two words which had and had not been paired during list presentation. Correct responses were, of course, yes and no, respectively.

The serial position of the critical pair (i.e., that pair tested in the recognition test) was defined in terms of the first (or A) member of the A-B pair. With 54 lists on four sessions there was a total of 216 trials, and each of the six serials positions (SP) was tested equally often. For each serial position half the recognition tests were A-B ("signal") while half were A-X ("noise"). For Serial Positions 1 and 6 the X terms were always the second (or B) members of the pairs from Serial Positions 2 and 5, respectively. For Serial Positions 2–5 the X terms came equally often from the immediately preceding and succeeding pair (though always the B member of the adjacent pair).

All pairs consisted of common English words taken from the Thorndike-Lorge (1944) list with G count of 20 or above. All homonyms, contractions, archaic words, and words with more than eight letters were excluded. Words were paired at random and without restriction. Pairs were photographed on 16-mm. film and projected by a single-frame Dunning animatic projector. The two words in each pair were displayed vertically rather than horizontally (i.e., one above the other rather than side by side). A visual noise field was inserted between each pair; this visual noise field consisted of letters of the alphabet typed at random and as densely as possible. Thus, on the filmstrips A-B pairs and the visual noise field alternated regularly in an ababa order.

There were four different filmstrips, one for each session, and their order of presentation was counterbalanced across *S*s in a 4 × 4 Latin square. Also, orthogonal to this counterbalancing A-B and A-X pairs were reversed for half the *S*s. That is, any pair which had been tested as A-B for half the *S*s was tested as A-X for the other half, and vice versa. Within each of the filmstrips the order of presentation of conditions was randomized in blocks of 12 trials (i.e., six serial positions and A-B or A-X test trials), and this randomized order naturally differed on the four filmstrips.

Stimulus presentation was programed by Hunter timers; A-B pairs and visual noise fields alternated at a 1-sec. rate. Each list presentation began and ended with a visual noise field. At the end of each list (i.e., when the visual noise field following the sixth pair had appeared) *S* was handed an index card with two words written on it, one above the other. He then wrote down on his data sheet either "yes" or "no"; yes if it matched one of the pairs in the list and no if it did not. Also, he had to write down a number from 1 to 5 indicating how confident he was that he was correct. In the instructions *S* was told, "1 is a pure guess, 3 is average certainty, 5 if you are very sure, and 2 and 4 are degrees between to help you rate." Also, a description of this rating scale was constantly in view for *S* to refer to throughout the experiment.

There were 24 *S*s, 12 students of each sex who were fulfilling a requirement of the introductory psychology course. On the first day there were two practice trials, but none thereafter. After each list presentation there was a 15-sec. interval for *S* to respond, and he was instructed to make a choice on every trial. The end of the 15-sec. interval was indicated by an audible click from one of the timers, and a "ready" signal preceded the start of the next list by several seconds.

RESULTS

The probability of a correct recognition (i.e., yes for A-B and no for A-X) was .764, .745, .720, .730, .874, and .927 for Serial Positions 1–6, respectively. An analysis of variance (of the number of correct recognitions) showed that the effects of serial position were highly significant; $F(5, 115) = 27.55$, $p < .001$. However, individual *t* tests showed that, at the .05 level, none of the comparisons involving the first four serial positions differed significantly from one another. Therefore, SP 1–4 were combined for subsequent analysis.

The ROC curves for SP 1–4 and for SP 5 are shown in Fig. 1. The curves were constructed following the method described by Pollack and Decker (1958). To indicate this method, the raw data for SP 1–4 are shown in Table 1. With four serial positions, 18 replications, and 24 *S*s, there were a total of 1,728 trials for the A-B and for the A-X presentations. The first two rows of Table 1 show the distributions of these 1,728 trials over the 10 possible confidence judgments. The confidence judgments are ordered left to right from "No-5" to "Yes-5" to correspond with a criterion ordering from lax to strict, respectively. The conditional probabilities, shown in the bottom two rows of Table 1, are cumulative, the probability that a given confidence rating was assigned to Category *j or stricter*. Thus, one starts from the right and moves to the left, separately for A-B and A-X.

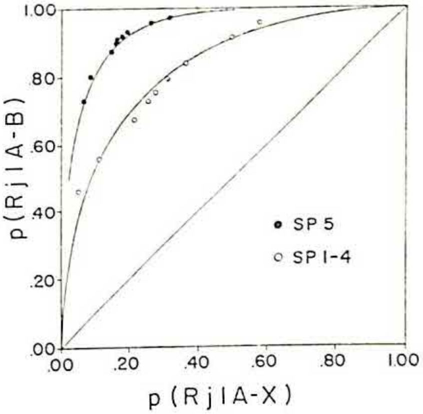


FIG. 1. Cumulative conditional probabilities of ratings R_j or stricter for A-B and A-X presentations.

The typical ROC curve shows hits (ordinate) as a function of false alarms (abscissa). Here, the cumulative

TABLE 1

DISTRIBUTION OF CONFIDENCE JUDGMENTS AND CONDITIONAL PROBABILITIES FOR A-B AND A-X PRESENTATIONS (SUMMED OVER FOUR SERIAL POSITIONS, 18 REPLICATIONS, AND 20 *S*s)

	No					Yes					Total
	5	4	3	2	1	1	2	3	4	5	
A-B	75	75	128	80	65	49	93	203	165	795	1728
A-X	732	137	231	94	58	37	73	172	102	92	1728
$P(R_j \mid \text{A-B})$		.957	.913	.839	.793	.755	.727	.673	.556	.460	
$P(R_j \mid \text{A-X})$		.576	.497	.363	.309	.275	.254	.212	.112	.053	

conditional probability of assignment to Category *j* or stricter is shown for A-B presentations (ordinate) as a function of A-X presentations (abscissa). The rationale behind this method has been given by Egan (1958); Egan, Schulman, and Greenberg (1959); and Pollack and Decker (1958). Basically, the idea is that *S* is capable of adapting multiple criteria, and a rating method provides more information about these criteria than one obtains from a binary (yes-no) judgment. In the present case, a judgment of "yes-5" represents the strictest criterion; "yes-4" is next most strict, and any observation that exceeded the criterion in the former case would also do so in the latter. Then, with 10 ratings, there are nine criteria ranging, as indicated, from strict to lax.

The data points for SP 1–4 and for SP 5 are well fitted by the smooth curves of signal-detection theory. (Data for SP 6 are not shown because the confidence ratings were so extreme.) In order to construct these smooth curves the data points were first plotted on normal-normal coordinates (Swets, Tanner, & Birdsall, 1961) and the best-fitting straight line drawn by eye. For SP 1–4 the slope of this line was close enough to unity to accept the assumption of equal variances for the A-B and the A-X distributions, but for SP 5 the slope was approximately 1.25. The obtained values for d' were 1.36 and 2.0 for SP 1–4 and SP 5, respectively. The difference in symmetry between the two curves simply reflects the fact that the equal-variance assumption held only for SP 1–4.

As has been pointed out elsewhere (e.g., Egan, 1958; Swets, 1961), the high-threshold assumption predicts that the data points should lie along a straight line running from the left-hand vertical axis to the upper right-hand corner of the graph. The *y* intercept would indicate the "true" threshold and the slope would be equal to one minus this threshold proportion. The linear function is a consequence of the fact that, according to a high-threshold assumption, *O* is incapable of ordering his judgments below the threshold. Thus, an increase in the hit rate can only be achieved by more false alarms, and these must increase at the same rate. As can be seen from Fig. 1, the data points do not fall along a straight line, and it would seem that the results do not support the high-threshold assumption.

DISCUSSION

From the results of the present experiment it does not seem correct to postulate a high threshold in short-term memory (STM), above which one can make correct recognitions and below which one responds by chance. Instead, recognition in STM seems more like the continuous process postulated by signal-detection theory with the individual capable of varying his criterion over a wide range. Of course, probably the data could also be accounted for by various low-threshold models as well (see Norman, 1964); however, the present experiment is too crude to make such a test very feasible.

There are, of course, various objections that could be raised about the present study and its conclusions. First, it could be argued that signal-detection theory has nothing to do with paired associates, or that ROC curves have no place in STM. Admittedly, one cannot specify the waveform or bandwidth of an associative bond, but signal-detection theory provides a model not only for the sensory processes but also for *S*'s decision processes. And, if a threshold is invoked as an explanatory concept, it seems a little unfair to deny the possibility of an empirical test.

A second objection is that the postulated threshold applies to recall, not to recognition. Thus, as used by Underwood and Keppel (1962), the high-threshold concept is applied to studies which test STM by recall. However, by now we have collected sufficient data with paired associates so that we can say with a fair degree of confidence that about 25% of the items from SP 1–4 and about 60% of the items from SP 5 would have been recalled correctly had retention been tested by recall (see Murdock, 1963, page 438, Table 5). Therefore, in terms of the argument of Underwood and Keppel (1962), some of the pairs would have been above and some below their postulated "anticipatory threshold" (page 4, Fig. 2). Recognition rather than recall was used in the present experiment simply because it is generally considered to be a more sensitive test of retention.

A third objection is that the results obtained are an artifact of averaging. That is, in many signal-detection studies one obtains intensive data from a few *S*s and plots ROC curves individually; here the data from all 24 *S*s were pooled to obtain the overall curves of Fig. 1. However, in a study of recognition memory for common English words Egan (1958) found that averaging did not appreciably modify the ROC curves for individual *S*s. And, it might be noted, he too found little evidence for a high-threshold concept. In the present experiment we did try to plot ROC curves for individual *S*s, but the data were too variable to be very revealing.

In conclusion, it would seem that the methods of signal-detection theory may provide a useful approach to the study of problems in STM. This theory, or model, is well developed and leads to a number of testable deductions, and its quantitative nature may be a useful antidote to the vague theorizing that is sometimes characteristic of the verbal-learning field. While it may be too much to think of measuring the strength of memory traces in terms of d' and β, still it should be profitable to apply some of the basic concepts of signal-detection theory to short-term memory.

REFERENCES

BROADBENT, D. E., & GREGORY, M. Division of attention and the decision theory of signal detection. *Proc. Roy. Soc. London, Ser. B*, 1963, **158**, 222–231.

EGAN, J. P. Recognition memory and the operating characteristic. Technical Note AFCRC-TN-58-51, 1958, Indiana University, Hearing and Communication Laboratory.

EGAN, J. P., SCHULMAN, A. I., & GREENBERG, G. Z. Operating characteristics determined by binary decisions and by ratings. *J. Acoust. Soc. Amer.*, 1959, **31**, 768–773.

ESTES, W. K. Learning theory. *Annu. Rev. Psychol.*, 1962, **13**, 107–144.

MURDOCK, B. B., JR. Short-term retention of single paired associates. *J. exp. Psychol.*, 1963, **65**, 433–443.

NORMAN, D. A. Sensory thresholds, response biases, and the neural quantum theory. *J. math. Psychol.*, 1964, **1**, 88–120.

POLLACK, I., & DECKER, L. R. Confidence ratings, message reception, and the receiver operating characteristic. *J. Acoust. Soc. Amer.*, 1958, **30**, 286–292.

SWETS, J. A. Is there a sensory threshold? *Science*, 1961, **134**, 168–177.

SWETS, J. A., TANNER, W. P., & BIRDSALL, T. G. Decision processes in perception. *Psychol. Rev.*, 1961, **68**, 301–340.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, Bureau of Publications, 1944.

UNDERWOOD, B. J., & KEPPEL, G. One-trial learning? *J. verbal Learn. verbal Behav.*, 1962, **1**, 1–13.

(Received July 8, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 5, 448–451

INTERACTION OF TIME-UNCERTAINTY AND RELATIVE SIGNAL FREQUENCY IN DETERMINING CHOICE REACTION TIME[1]

PAUL BERTELSON AND JACQUES BARZEELE

Université libre de Bruxelles

Choice RTs to 2 signals of relative frequencies .2 and .8 were measured under 2 time-uncertainty conditions: constant FPs of either .5 or 5.0 sec. The RT to the more frequent signal is more affected by time uncertainty than the RT to the less frequent one. This result shows that "preparation," a hypothetical state which has often been postulated to account for time-uncertainty effects, is at least in part specific to one particular signal-response pair, or, alternatively, that the well-known effect of signal relative frequency is affected by very short-term fluctuations.

It is now well-established that both simple (Botwinick & Brinley, 1962; Karlin, 1959; Klemmer, 1956, 1957; Woodrow, 1914) and choice reaction time (Bertelson & Borsu, 1965; Boons & Bertelson, 1961) are influenced by the duration and the variability of the foreperiod (FP). This effect can be accounted for by a simple model already suggested by Woodrow. Fast responding depends on the existence of a state of high responsiveness, which has been called "readiness," "attention" (Woodrow, 1914), "preparatory set," or "expectancy" (Mowrer, 1940), or, as it will be here, "preparation." Preparation is of very short duration, so that it is important that it is present at the right time. Failures to synchronize preparation with stimulus onset can result either from the impossibility of predicting when the signal will come (as with variable FPs), or from the inaccuracy of internal time (as with constant but long FPs). Klemmer (1956) introduced the term "time uncertainty" to express the combination of both factors.

One question which can be asked about preparation is whether it is specific to one particular signal-response pair, i.e., leads to shorter RT only if the particular signal which is expected, occurs. In a former paper (Bertelson & Boons, 1960), the authors held that the effect of time uncertainty on *choice* RT by itself proves that some kind of nonspecific preparation is possible. It was later realized (Boons & Bertelson, 1961) that the effect is compatible with the hypothesis of strictly selective preparation: With low time uncertainty, preparation for one of the signals would always be achieved at the right time, permitting a gain of time whenever that signal occurs; whereas with high time uncertainty, there would sometimes be no preparation at all, the stimulus arriving in the interval between two preparation phases, or while *S* is attending to some different source of stimulation.

The previous experiments concerning the effect of time uncertainty on choice RT were all done with equiprobable signals. In that situation, it is impossible to discriminate between the two hypotheses, because one has no information about preparation achieved on a particular trial,

---

[1] This work was partly supported by a "Crédit aux Chercheurs" from the Belgian Fonds National de la Recherche Scientifique. Thanks are due to J. A. Leonard and L. H. Shaffer for very useful suggestions.

except by inference from the observed RT. If one considers a situation with signals of unequal frequencies, one may postulate that if preparation were specific, *S* would prepare for frequent signals more often than for less frequent ones. Hence, when preparation is made difficult by time uncertainty, the RT to a frequent signal would be more affected than the RT to a less frequent one.

In situations involving a frequency bias, a selective preparatory process has often been assumed to account for the well-known effect of frequency on RT (Falmagne, 1965; Fitts, Peterson, & Wolpe, 1963; Hyman, 1953; Leont'ev & Krinchik, 1963). But we have no right to identify this process with the preparation phenomenon revealed by time-uncertainty manipulations. The former is selective by definition, but has not been proved so far to be short term, just as the latter is short term by definition but has not been proved to be selective. We have thus to consider two possibilities. One is that two *separate* preparatory processes exist, one specific and permanent (for the duration of one trial), the other nonspecific and short term. The first process would give shorter RTs on all trials where specific preparation existed for the signal which occurred, whatever its time of occurrence. The second would give shorter RTs whenever short-term preparation existed at the time of occurrence of the signal, whatever the nature of that signal. In that case, provided the frequency of preparation to particular signals does not change with the FP, we should expect to find no interaction between time uncertainty and signal relative frequency. The other possibility is that there exists one process, which is at the same time specific and short term. It would give shorter RTs only when the particular signal which was expected arrived while good preparation was present. Ill-timed preparation for the right signal or well-timed preparation for the wrong signal would both be ineffective. Under that hypothesis, we should expect to find an interaction between time uncertainty and relative signal frequency.

For the sake of simplicity, specificity has been considered as an all-or-none affair: preparation affects the RT to one signal and has no effect on those to the other signals. This assumption might be too extreme, and one has to consider the possibility of relative specificity, or specificity with generalization. Specific preparation would now affect the RTs to several, or even to all signals, but by different amounts. If this weaker meaning is adopted, the predictions are still the same: interaction between time uncertainty and signal frequency will be observed only if specific preparation is not constant over the duration of one trial.

The experiment to be presented consisted of measuring choice RTs to two signals of unequal frequencies, with constant FPs of either .5 or 5.0 sec. A constant FP of $\frac{1}{2}$ sec. is known to give about optimal warning (Boons & Bertelson, 1961; Karlin, 1959), and thus constituted the low time-uncertainty condition. For the high time-uncertainty condition, a long constant FP was chosen rather than variable ones because (*a*) it was technically easier to obtain, and (*b*) the literature suggests that variable FPs can induce in *S*s a range of different strategies, depending on as yet little understood conditions.

METHOD

Apparatus.—The display was constituted by a Nixie numerical indicator (a cold cathode tube with figure-shaped cathodes) fixed behind a red glass (to reduce brightness) at eye level, on a vertical panel, at about 1 m. from *S*, and a neon pea lamp situated 2 cm. below the Nixie. On each trial, the latter first gave a brief flash (about 50 msec.), which constituted the warning signal. After an interval constituting the FP, either the digit 1 or the 2 appeared on the Nixie. The *S*, seated in armchair, responded by pushing one of two keys, fixed one to each armrest, in a vertical plane. They consisted of flexible strips of steel, resting on microswitches. The *S*s were instructed to use the S-R correspondence: 1-left, 2-right. A constant interval of 10 sec. intervened between each reaction signal and the following warning signal. The time intervals were controlled by a timing unit based on capacitance-resistance circuits. The sequence of reaction signals was controlled by a punched tape reader. The RTs to the two signals were cumulated on two separate chronoscopes. They were at the same time recorded on paper by a pen recorder running at 5 cm/sec which permitted readings to the nearest .01 sec. Errors were recorded on electromagnetic counters.

Subjects.—Ten volunteer students, aged 18–20, four boys and six girls, participated in three sessions each.

Procedure.—One run consisted of 100 trials. Signal 2 was presented on 20 trials, chosen at random without restriction for each of 20 different program tapes, and Signal 1 on the other 80 ones. On each session, two runs were given, one with each of the two FP durations of .5 sec. and 5.0 sec. One FP condition was given first to one half of the *S*s and the order was reversed from one experimental session to the other. The first session was considered as a practice one and discarded from the analysis. The *S*s were informed verbally of the relative frequencies of the two signals. They were told the overall mean RT and the number of errors at the end of each run of 100 reactions.

RESULTS

Table 1 shows that the mean RT to the frequent signal is increased by 65 msec. when the FP is increased from .5 to 5.0 sec., against only 31 msec. in the case of the infrequent signal. This pattern is observed at the level of individual means in 9 *S*s out of 10. By a two-tailed sign test, the hypothesis that the RTs to both signals are equally affected by time uncertainty can be rejected at $p = .021$. Subject 2 shows a slightly larger effect of FP duration on the less frequent signal. A fourth session was conducted, and the larger effect was now observed for the more frequent signal.

TABLE 1

MEAN REACTION TIME (IN MSEC.) FOR SESSIONS 2 AND 3

*S*s	Foreperiod (in Seconds)					
	Signal 1 ($p = .8$)			Signal 2 ($p = .2$)		
	.5	5.0	Diff.	.5	5.0	Diff.
1	340	449	109	453	471	19
2	279	343	64	358	430	72
3	286	324	38	374	388	14
4	333	401	68	430	439	9
5	321	382	61	344	394	50
6	396	438	44	471	479	8
7	291	319	28	344	338	−4
8	330	354	24	370	371	1
9	391	488	97	457	517	60
10	347	467	120	440	524	84
M	331	396	65	404	435	31

Errors occurred at the same low frequency (1.5%) with the two FPs. They are more frequently made to the infrequent signal (2.7%) than to the other one. The sample is too small to answer questions about possible interactions.

Although the predictions to be tested were about mean RT, it was felt interesting to look for possible changes in the shapes of the RT distributions. For that purpose, a random sample of 20 reactions to the .8 probability signal was taken for each *S* under each FP condition at each session, and was analyzed together with the 20 reactions to the .2 probability signal. For each sample, RTs occupying Ranks 1, 3, 6, 9, 10, 15, 18, and 20 were tabulated, and averaged across *S*s and sessions to give the mean cumulative fre-

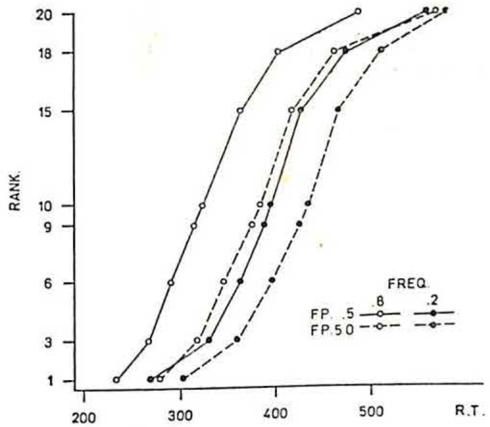


FIG. 1. Mean cumulative frequency distributions of RTs. (Ordinate: rank in a sample of 20 RTs, when ordered according to length; Abscissa: mean RT—for 10 *S*s, two sessions each—at that rank, in milliseconds.)

quency curves appearing on Fig. 1. One observes a larger effect of FP duration on the RT to the more frequent signal at all cutting points. Neither time uncertainty nor signal frequency seem to affect one region of the RT distribution rather than another.

DISCUSSION

The effect of time uncertainty clearly interacts with that of signal frequency. This result favors the hypothesis of a preparatory process at the same time selective and of short duration.

REFERENCES

BERTELSON, P., & BOONS, J. P. Time-uncertainty and choice reaction time. *Nature*, 1960, **187**, 531–532.

BERTELSON, P., & BORSU, M. Incertitude temporelle et reconnaissance tachistoscopique. *Année psychol.*, 1965, in press.

BOONS, J. P., & BERTELSON, P. L'influence de l'incertitude temporelle sur le temps de réaction de choix. *Année psychol.*, 1961, **61**, 361–376.

BOTWINICK, J., & BRINLEY, J. F. An analysis of set in relation to reaction time. *J. exp. Psychol.*, 1962, **63**, 568–574.

FALMAGNE, J. C. Stochastic models for choice reaction time with applications to experimental results. *J. math. Psychol.*, 1965, **12**, 77–124.

FITTS, P. M., PETERSON, J. R., & WOLPE, G. Cognitive aspects of information processing. II: Adjustments to stimulus redundancy. *J. exp. Psychol.*, 1963, **65**, 507–514.

HYMAN, R. Stimulus information as a determinant of reaction time. *J. exp. Psychol.*, 1953, **45**, 188–196.

KARLIN, L. Reaction time as a function of foreperiod duration and variability. *J. exp. Psychol.*, 1959, **58**, 185–191.

KLEMMER, E. T. Time uncertainty in simple reaction time. *J. exp. Psychol.*, 1956, **51**, 179–184.

KLEMMER, E. T. Simple reaction time as a function of time uncertainty. *J. exp. Psychol.*, 1956, **54**, 195–200.

LEONT'EV, A. N., & KRINCHIK, E. P. On certain aspects of human information processing. *Soviet Psychol. Psychiat.*, 1963, **1**, 24–31.

MOWRER, O. H. Preparatory set (expectancy)—Some methods of measurement. *Psychol. Monogr.*, 1940, **52**(2, Whole No. 233).

WOODROW, H. The measurement of attention. *Psychol. Monogr.*, 1914, **17**(5, Whole No. 76).

(Received June 22, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 5, 452–458

DISCRIMINATION LEARNING AS A FUNCTION OF VARYING PAIRS OF SUCROSE REWARDS[1]

ROGER W. BLACK

University of Iowa

5 groups of rats were trained in a Y maze using a combination of free- and forced-choice trials to equate number of Rs to each arm. Sucrose reward was present in both arms, but in different concentrations. In terms of the solutions received, the groups were: 2%–4%, 4%–8%, 8%–16%, 16%–32%, 32%–64%. Principal results were: discrimination occurred in both choice and starting-speed measures and the degree to which Ss discriminated in one measure correlated significantly with degree of discrimination in the other; running speed appeared to be a negatively accelerated, exponential function of sucrose concentration; optimal discrimination occurred for intermediate pairs of concentrations. These results were in accord with implications derived from Spence's interpretation of incentive motivation.

In his interpretation of the role of reinforcement in appetitive learning, Spence (1956) assumed that a primary function of reward is the evocation of the consummatory response (R_G). Because of its repeated pairing with cues in the goal box during the acquisition of an instrumental response, implicit components of the consummatory response (r_g's) become conditioned to the goal box and generalize to the runway and start box. The occurrence in the alley of r_g, along with its afferent component s_g, is assumed to contribute incrementally to the motivational level of *S* through the molar incentive-motivation variable, *K*. *K* combines additively with drive and multiplicatively with habit strength to produce reaction potential—i.e., $E = H \times (D + K)$.

The value of *K* depends upon the strength with which $r_g - s_g$ is evoked which, on any trial, depends in turn upon at least two variables. First, since $r_g - s_g$ is the result of classical conditioning, *K* will increase with number of rewarded trials, until the asymptotic strength of $r_g - s_g$ is reached. Second, the strength of $r_g - s_g$ depends upon the vigor of the overt consummatory response (R_G) which is assumed to be an increasing function of magnitude of reward. Specifically, when a sufficient number of rewarded trials have been provided to insure that the conditioning of $r_g - s_g$ is asymptotic, the value of *K* is given by

$$K = M(1\text{–}10^{-aR}) \qquad [1]$$

where

M = maximum possible value of *K*.
R = magnitude of reward.
a = rate of growth parameter.

Now if two response tendencies, based on equally strong habits and levels of drive but different reward magnitudes are considered, the difference in the strengths of these responses will depend upon the difference in the values of *K* associated with them. With the responses at their asymptotic strengths, this difference will be given by:

---

[1] This report is based on a dissertation submitted to the Department of Psychology of the University of Iowa in partial fulfillment of the requirements for the PhD degree. The writer is greatly indebted to Kenneth W. Spence under whose direction this investigation was conducted.

$$K_L - K_S = M(1\text{–}10^{aR}) - M(1\text{–}10^{-a1/nR}) \quad [2]$$

where K_L and K_S indicate the Ks resulting from the large (R_L) and small (R_s) rewards, respectively, and n is the factor by which R_L exceeds R_s.

Further, if the ratio (R_L/R_s) is held constant, as the absolute amount of reward increases from some very small value, the difference ($K_L - K_S$) first increases, reaches a maximal value, and then decreases. In general, this difference will be maximal when R_L has a magnitude given by:

$$R_L = \frac{n \log n}{a(n - 1)},$$

where a and n are the quantities of Equations 1 and 2. The derivation of this expression is analogous to one described by Spence (1956, pp. 244–245).

From these considerations, several predictions may be made with respect to performance in a discrimination situation (e.g., a T maze) in which (*a*) the alternative responses are *both* reinforced but with different reward magnitudes; (*b*) *S*s receive pairs of rewards which differ in mean magnitude but are in the same (R_L/R_s) ratio; (*c*) the number of occurrences of each response and thus, according to Spence, their habit strengths, are equated. For such a situation it may be predicted that *S*s should respond more *quickly* to the positive (larger reward) than the negative arm and respond more *frequently* to that arm when choice is possible, since the value of K for that arm will be greater. Second, considering independent groups receiving pairs of rewards with successively larger mean magnitudes, average response speed should increase in a negatively accelerated, exponential fashion. Third, as mean reward magnitude increases, the degree to which discrimination occurs (in both speed and choice measures) should first increase and then decrease. Finally, the larger of the pair of reward magnitudes (i.e., R_L) for which discrimination is maximal should be given by Equation 3, which is solvable if an estimate of the rate of growth parameter, a, can be obtained.

METHOD

Experiment I

Subjects.—The *S*s were 40 female, hooded rats, 85–105 days old at the beginning of training, from the colony maintained by the Department of Psychology of the University of Iowa.

Apparatus.—The apparatus included an enclosed Y maze, together with electronic circuitry to provide response-time measures. The maze consisted of an 18-in. long retaining box, a 12-in. start box attached to the retaining box, and two 30-in. arms which coverged upon the start box to form a Y. The maze was 12 in. high, constructed of wood painted black, and covered with a glass top. The start box and arms were 2 in., and the retaining box 3.5 in., wide. At the end of each arm was a wood block, also painted black, with a metal-lined indentation at its center which served as the goal cup.

The retaining box was separated from the start box by a guillotine door while a second guillotine door at the end of the start box opened onto the choice point and served as the start door. A retrace door was located within each arm of the maze at a point 9 in. from the start door. These doors were so constructed that they could be used to hold a block of wood which, when used, fitted flush with the wall of the start box and extended that wall to the distal end of the choice point, thus completely covering the entrance to that arm.

Photoelectric circuitry, together with Hunter klockounters, provided starting- and running-time measures. Start time was measured from the opening of the start door until *S* reached a point within either arm 12 in. beyond the start door. Running time was measured from that point to a point 12 in. farther in the arm.

Procedure.—Two weeks prior to the beginning of training, all *S*s were placed on a 23-hr. food-deprivation schedule with water continuously available.

The *S*s were randomly divided into five groups of eight *S*s each. Defined in terms of the sucrose solutions received during training, the groups were: 2%–4%; 4%–8%; 8%–16%; 16%–32%; 32%–64%. Thus, the members of each pair of solutions were in 2:1 ratio. Solutions were prepared from ordinary cane sugar and distilled water and are stated in percentage of sucrose by weight.

For each *S*, one arm of the maze was designated "positive" and always contained the sweeter solution, the "negative" arm always containing the less sweet solution. For half of the *S*s, the right, and for half the left, arm was positive. Six acquisition trials were given daily for 20 days, the first trial of each day being a free-choice trial and the remaining forced trials (i.e., one arm blocked). Forced trials were so administered that *S* responded to each arm three times daily. Specifically, if the initial choice was the positive (P) arm, the remaining trials were administered according to one of the following sequences: PNNPN; NPNNP. For a negative (N) initial choice, the sequences were: PPNNP; PNPPN. Following the choice trial on a given day, *S* was assigned to the sequence which was the alternate of the one to which it was assigned on the previous day it had made the same choice.

On choice trials, the procedure was as follows: .5 ml. of the appropriate sucrose solution was placed in the goal dish in each arm; *S* was placed in the retaining box and in 10 sec. the door to the start box was opened until *S* entered the start box, whereupon that door was closed and the start door was immediately opened. When *S* entered either arm, the retrace door was closed and *S* was allowed 10 sec. to consume the solution. On forced trials the procedure was the same except that one of the arms was blocked. The *S*s were run sequentially in squads of 10 which provided an intertrial interval of about 10 min.

Starting and running times were recorded only on the final two trials of a day, one of which was always a positive, and the other a negative, trial. The trial sequences employed, it should be noted, counterbalanced the serial position of the positive and negative trials on the final two trials, as well as the nature of the trial preceeding the final two trials.

Experiment II

A difficulty developed in the first experiment in that closing the retaining box door as *S* entered the start box resulted in many of the *S*s orienting away from the start door and toward the retaining-box door. Since these *S*s often spent considerable periods of time before reorienting toward the choice point, and since the start-time clock was operating during this period, a large number of extremely inflated "starting times" were recorded. This, coupled with a difficulty in the photoelectric circuit which resulted in the loss of a substantial number of additional starting times, rendered the starting-time data of the first experiment essentially useless. Experiment II was a replication of the first experiment except that these sources of difficulty were eliminated.

Subjects.—The *S*s were 50 female, hooded rats, 80–92 days old at the beginning of training, from the colony maintained by the Department of Psychology of the University of Iowa.

Apparatus.—The apparatus was identical to that in Exp. I except that the retaining box was removed.

Procedure.—The procedure was also identical to that of Exp. I, except that *S* was placed directly into the start box. The start door was opened 5 sec. thereafter, provided that *S* was oriented toward the start door. Ten *S*s were assigned to each of the five groups.

RESULTS AND DISCUSSION

Starting speed.—Acquisition curves for the second experiment indicated that in each group starting speed to the positive arm approached a higher asymptote than that for the negative arm. Nevertheless, the actual differences between the positive and negative curves for the various groups were generally rather small and suggested that discrimination between the sucrose solutions was difficult. It further appeared that average starting speed increased as mean sucrose concentration increased. To evaluate these effects, each *S*'s mean starting speed on Trials 90–120 was determined. An analysis of these "asymptotic speed scores" indicated that overall start speed (i.e., ignoring "arms") was an increasing function of mean sucrose concentration, $F\ (4,\ 45) = 6.86,\ p < .001$.

Thus, mean asymptotic start speed was 1.16 ft/sec for Group 2%–4% and 1.46 ft/sec for Group 32%–64%. When differences in start speed to the two arms were considered, reliable discrimination was also found to occur, $F\ (1,\ 45) = 30.72,\ p < .001$. Figure 1 indicates the difference in asymptotic mean starting speeds to the arms (positive minus negative) for the five groups. This difference was minimal for the 2%–4% group, rose to a maximum for the 16%–32% group, and then decreased for the 32%–64% group. While this finding confirmed the predictions stated earlier, when tested as the Arms × Groups interaction, it failed to prove statistically reliable, $F\ (4,\ 45) = 1.57$.

Choice data.—Figure 2 indicates percentage of choice of the positive arm for the 90 combined *S*s of Exp. I and II. The similarity of this function to that in Fig. 1 is apparent in that both measures show least discrimination for Group 2%–4%, maximal discrimination at an intermediate value of sucrose concentration, and a subsequent decrement in discrimination. The two functions, however, reached their maxima at somewhat different values. Analysis of variance indicated that the differences between groups in asymptotic percentage of choice were reliable, $F\ (4,\ 85) = 2.78,\ p < .05$. Thus, for both starting speed and percentage of choice, the predicted general relationship between degree of discrimination and mean sucrose concentration was obtained. The relationship itself, however, was statistically confirmed only for the choice data.

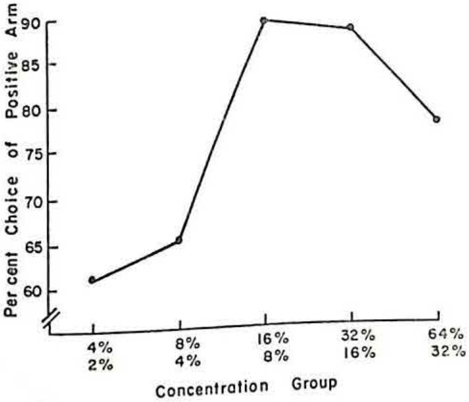


FIG. 1. Difference in starting speed to the positive vs. the negative arm as a function of concentration group.

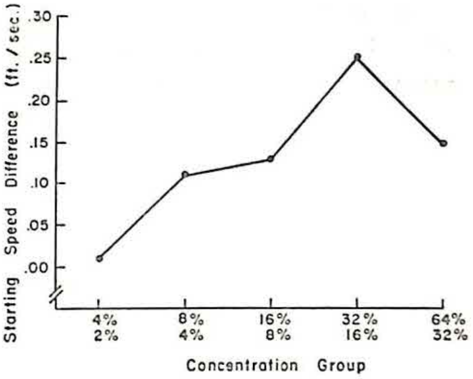


FIG. 2. Percentage of choice of the positive arm as a function of concentration group.

The general correlation between discrimination in terms of speed differences and in terms of percentage of choice which is evident in the similarity of Fig. 1 and 2 is required by a "macromolar" theory such as Spence's. It is, however, not required by the "micromolar" approach advanced by Logan (1956). A macromolar interpretation assumes that differences in the speed with which *S*s respond to the arms of a maze and the percentage of times *S* responds to each arm on choice trials are both reflections of the difference in the same underlying reaction potentials. Thus, these measures—speed differences and percentage of choice—must be positively correlated. Further, in the case of the Hull-Spence theory the correlation, when the choice data are appropriately scaled, must be linear (Logan, 1953).

In order to test this aspect of the theory, the mean difference in start speed to the positive vs. negative arm was computed for *S*s in Exp. II for Trials 60–90. Percentage of choice of the positive arm was also computed for these trials and these scores were then converted to normal deviates, according to a procedure described by Logan (1953). (Percentage of choice scores of 100% or 0% were omitted because they cannot legitimately be scaled.) For the present data, the correlation between these measures was, $r = .45$, Eta $= .60$, and analysis of variance indicated that the linear regression component of this relation was significant, $F\ (1,\ 48) = 8.58$, $p < .01$, while the deviation from linearity was not, $F\ (4,48) = 2.74$. This finding, which was recently replicated by Fowler (1963), renders the macromolar approach tenable for the present data. It does not, of course, constitute a "critical" test of the micromolar-macromolar distinction since the micromolar position would also predict the obtained correlation between speed differences and percentage of choice under the present experimental conditions.

Running speed.—A comparison of the running-speed curves for the two experiments suggested that the sets of data were comparable. Moreover, statistical analysis performed on asymptotic running speeds (Trials 90–120) failed to indicate a significant difference between the experiments or a significant interaction of "experiments" with any other variable (all p's $> .10$). Further, no discrimination between the arms occurred in terms of running speed. Thus, *S*s ran about equally quickly in the positive and negative arms, $F\ (1,\ 80) = 1.71$. That differential responding was observed only for starting speeds probably resulted from the fact that cues permitting discrimination were present to any degree only at the choice point— a portion of the alley included in the starting- but not the running-speed measures. Thus, at the choice point, *S* encountered differential cues on positive and negative trials of both a visual (i.e., which arm was "open") and a kinesthetic (direction of turn) nature. Beyond the choice point, however, visual cues were essentially identical while differential kinesthetic aftereffects of the turning response were also probably minor, the degree of turning required in the Y maze being slight.

Figure 3 indicates mean asymptotic running speed for the five groups as a function of sucrose concentration. Each group is designated on the abscissa in terms of the sweeter of the two solutions it received. As this figure indicates, performance increased as a negatively accelerated, monotonic function of sucrose concentration and an analysis of variance indicated that differences between the groups were reliable, $F\ (4,\ 80) = 29.94$, $p < .001$. To test

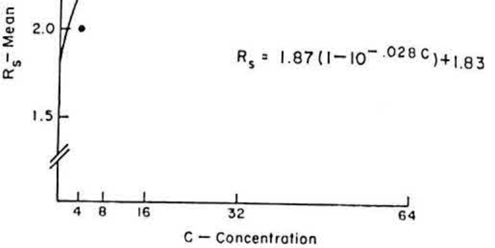


FIG. 3. Mean running speed as a function of sucrose concentration. (The abscissa specifies groups by the sweeter of the solutions they received.)

the prediction stated earlier, a curve of the form given in Equation 1 was fitted to these data using a least-square procedure described by Lewis (1960, pp. 76–80). This function, represented in Fig. 3, appeared to approximate the data quite well. In addition, it provided an estimate of the rate parameter, a, of Equations 1–3—i.e., $a = .028$. Entering this value in Equation 3, and noting that the ratio of the large (R_L) to the small (R_s) reward (n) was 2, solution of Equation 3 indicated that maximal discrimination is predicted for a group receiving an 11%–22% pair of sucrose concentrations. While the present study did not employ an 11%–22% group, it did provide two estimates of the point of maximal discrimination—i.e., in terms of starting speeds and in terms of percentage of choice. Those estimates were 8%–16% and 16%–32% which fell to either side of 11%–22% and were as close to that value as the present selection of groups permitted.

While the function in Fig. 3 relating response strength to sucrose concentration is monotonic, other investigators have reported a decrement in response strength at high sucrose concentrations. Thus, in a free-responding situation, maximal response rate has been reported for sucrose concentrations of 16% (Guttman, 1953) and 20.1% (Guttman, 1954). It is possible, however, that the decrement in response strength observed in the Skinner box with concentrated sucrose solutions may result from satiation rather than an intrinsically lower incentive value of more concentrated solutions (cf., Collier & Myers, 1961). Thus, Young and his associates (e.g., Young, 1945; Young & Greene, 1953) report that when *S*s are given long (24-hr.) exposures to sucrose solutions, both rate and amount of solution consumed were maximal with a 5% concentration; but, when the exposure period was brief, rate and amount of solution consumed increased monotonically with concentration. In the present discrete-trial situation, which involved the daily administration of only 3 ml. of solution over a 1-hr. period, satiation was presumably of less importance. On the other hand, the observed monotonic relationship may have resulted from the fact that *S*s were given experience with more than one sucrose concentration. Fowler (1963), in a situation similar to the present one, found that for *S*s trained with a single sucrose concentration, running speed was greater for a 16% than a 32% solution. He also reported, however, that when *S*s were trained with a pair of different concentrations (e.g., 16%–32%), they ran faster to the more concentrated solution. It is possible, therefore, that the relationship described by Fig. 3 would not be reproduced in an experiment involving groups which received only a single sucrose concentration.

The present results clearly support the interpretation of incentive motivation proposed by Spence (1956). Thus, both the general and quantitative implications of that theory were confirmed with respect to the form of the incentive function, the effect of increasing mean reward magnitude on discrimination, and the expected point of maximal discrimination between pairs of incentives In addition, support was provided for the tenability of the macromolar analysis of learning for situations comparable to that described above.

REFERENCES

COLLIER, G., & MYERS, L. The loci of reinforcement. *J. exp. Psychol.*, 1961, 61, 57–66.

FOWLER, R. L. *Magnitude and delay of reinforcement in spatial discrimination learning.* (Doctoral dissertation, University of Tennessee) Ann Arbor, Mich.: University Microfilms, 1963, No. 185.

GUTTMAN, N. Operant conditioning, extinction, and periodic reinforcement in relation to concentration of sucrose used as reinforcing agent. *J. exp. Psychol.*, 1953, 46, 213–224.

GUTTMAN, N. Equal reinforcement values for sucrose and glucose solutions compared with equal sweetness values. *J. comp. physiol. Psychol.*, 1954, **47**, 358–363.

LEWIS, D. *Quantitative methods.* New York: McGraw-Hill, 1960.

LOGAN, F. A. Three estimates of differential excitatory potential. *Psychol. Rev.*, 1953, **59**, 300–307.

LOGAN, F. A. A micromolar approach to behavior theory. *Psychol. Rev.*, 1956, **63**, 63–73.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

YOUNG, P. T. Studies of food preference, appetite and dietary habit. *Comp. Psychol. Monogr.*, 1945, **19**, 1–44.

YOUNG, P. T., & GREENE, J. T. Relative acceptibility of saccharine solutions as revealed by different methods. *J. comp. physiol. Psychol.*, 1953, **46**, 295–298.

(Received June 24, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 5, 459–463

DETECTION OF A VISUAL SIGNAL WITH LOW BACKGROUND NOISE: AN EXPERIMENTAL COMPARISON OF TWO THEORIES[1]

RAYMOND H. HOHLE

Institute of Child Behavior and Development, University of Iowa

Yes-no responses were obtained from each of 5 Ss under different conditions of induced response bias during 9 300-trial sessions (plus 9 extra sessions for 1 S) where a low-intensity stimulus was present or absent in a random sequence. Least-squares curves relating probabilities of a "yes" response with the stimulus present to probabilities of a "yes" with the stimulus absent were determined for each set of data, assuming (a) the theory of signal detection (TSD), and (b) the fixed-criterion hypothesis. The latter theory afforded a closer fit for each of the 6 sets of data when compared to the simplest form of TSD. A more general form of TSD led to curves which fit as well as those from the fixed-criterion theory, but this form has other, undesirable, implications.

One type of experiment which has served as a testing ground for signal detection theory involves presentation of a series of trials on which a low-intensity stimulus is present or absent in a random sequence. On each trial *S*'s task is to report "yes" or "no," depending on whether he thinks the stimulus was present or absent. From the data of such an experiment the theory of signal detection purports to extract a measure of *S*'s sensitivity which is independent of his response criterion. The usefulness of this measure, of course, depends on whether the theory provides reasonably accurate representations of *S*s' behavior.

The theory assumes that on every trial, whether the signal is present or not, *S* receives sensory input. For those trials on which no signal is presented this input consists only of randomly varying noise; and introduction of the signal simply adds a fixed level of input to the noise distribution. The *S*'s task on each trial is thus that of deciding whether his sensory input was from a distribution of noise alone or from a distribution of noise augmented by input from a signal. The basis for *S*'s decision may be seen from Fig. 1. The line, x, represents some monotonic transformation of sensory input level (*S*'s decision axis), and the two curves represent the probability density functions of x on the two kinds of trials (signal present or not present). If on a given trial, x takes a value greater than a critical value, c, *S*'s decision will be "yes, the signal was present"; if x was less than c, the decision will be "no, the signal was not present."

It is apparent from Fig. 1 that as *S*'s decision criterion is shifted, the conditional probabilities $\Pr\{\text{"yes"}|\text{Signal}\} \equiv P_S$, and $\Pr\{\text{"yes"}|\text{No Signal}\} \equiv P_N$ will vary together. If both the "Noise" and "Noise + Signal" distributions are assumed to be normal distributions with a common variance, and d is the difference between the means of the distributions (expressed in units of the common *SD*), then the curve relating the two conditional probabilities—*S*'s receiver operating characteristic, or ROC,

---

[1] The data on which this paper is based were reported at the Midwestern Psychological Association meetings at St. Louis, April 1964.

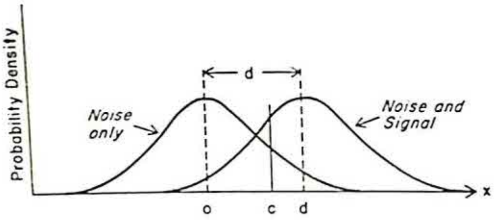


FIG. 1. Probability density functions of the variable representing *S*'s decision axis with no signal (noise only) and with a signal present.

curve—will be

$$P_S = 1 - F[F^{-1}(1 - P_N) - d], \quad [1]$$

where F is the unit normal cumulative distribution function, and F^{-1} is its inverse.

Equation 1 has been shown to provide reasonably good representations of empirical ROC curves for auditory stimuli (Swets, 1961b), and accordingly, fitted values of d have been interpreted as relatively pure measures of *O* sensitivity (e.g., Swets & Sewall, 1963).

A favored procedure in experiments designed to provide empirical tests of Equation 1 has been "to add enough noise . . . (a background of some kind) to bring the noise to a level where it can be measured [Swets, 1961a, p. 56]." One might raise the question, therefore, whether the theory of signal detection is applicable to traditional psychophysical experiments where an attempt is usually made to minimize stimulus noise external to *O*.

An alternate theory, originally proposed by H. R. Blackwell and W. M. Kincaid,[2] assumes that a signal is detected only when the hypothetical input variable exceeds some critical (threshold) level. This critical level is almost never exceeded when no signal is presented; nevertheless, under appropriate conditions (e.g., where high rewards are associated with detections), guessing on the part of *S* will produce "yes" reports on some trials even though the threshold value has not been exceeded. The structure of the Blackwell theory is represented in Fig. 2. On trials with the signal present a detection is made (input exceeds threshold) with probability P_D, the value of which depends on signal intensity and *S*'s sensitivity. In this event the response is always "yes." If there is no detection—which occurs with probability $1 - P_D$ on "signal" trials and nearly always on "no signal" trials—a guessing response "yes" will occur with probability P_G. The magnitude of P_G, of course, reflect *S*'s prevailing response bias, which depends in part on the particular experimental conditions.

The ROC curves implied by Blackwell's theory are straight lines passing through the points ($P_N = 0$, P_S

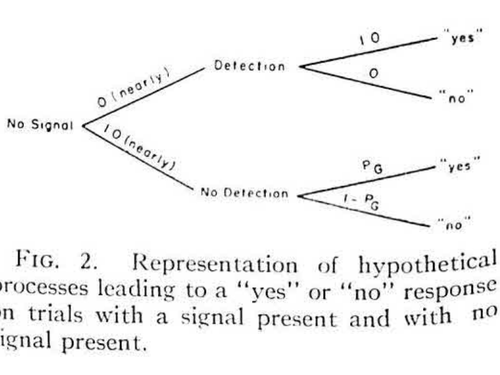


FIG. 2. Representation of hypothetical processes leading to a "yes" or "no" response on trials with a signal present and with no signal present.

---

[2] Colloquium address, University of Michigan, November 30, 1951. The theory has been referred to elsewhere (Swets, 1961b) as simply "the Blackwell threshold theory." For the sake of brevity, this practice will be followed here.

$= P_D$) and ($P_N = 1, P_S = 1$). From Fig. 2 we see that the conditional probability "yes," given a signal, is

$$P_S = (P_D)(1.0) + (1 - P_D)P_G,$$

and the conditional probability "yes," given no signal, is

$$P_N \cong (.0)(1.0) + (1.0)(P_G).$$

Since $P_N \cong P_G$, P_N may be substituted directly in the expression for P_S, so that

$$P_S \cong P_D + (1 - P_D)P_N. \quad [2]$$

Under the assumptions of the Blackwell theory, then, the parameter, P_D, is, like *d* in the theory of signal detection, independent of an *O*'s response biases.

One experiment (Swets, 1954), using a forced-choice procedure, yielded results which were inconsistent with the assumptions of the Blackwell theory but consistent with the theory of signal detection. On the other hand, data from two out of three experiments ostensibly replicating the Swets experiment (Kincaid & Hamilton, 1959) failed to provide evidence for rejecting the Blackwell theory. The present study is an attempt to compare the two theories directly by determining whether Equation 1 or 2 provides a better representation of several empirical ROC curves for detection of a visual signal with relatively low background noise.

PROCEDURE

A trial consisted of a shutter opening for 800 msec., exposing a 4-cm. circular screen with a luminance of 1.8 footlamberts (ftl.). The signal to be detected was a 1-cm. circle of light which was or was not added to the center of the screen in a random sequence over trials. The luminance increments serving as the signal ranged from .02 to .07 ftl. for different *S*s, but was constant throughout an experiment for any given *S*. Changes in *S*'s response bias were induced by varying the relative frequency with which the signal was present over trials (.33, .50, or .67), and by varying the patterns of rewards and punishments associated with the trial outcomes,

$$\begin{pmatrix} \text{Signal, "Yes"} & \text{No Signal, "Yes"} \\ \text{Signal, "No"} & \text{No Signal, "No"} \end{pmatrix}.$$

The "payoff" matrices used were

$$\begin{pmatrix} 1 & -5 \\ -4 & 4 \end{pmatrix}, \begin{pmatrix} 3 & -4 \\ -6 & 3 \end{pmatrix}, \text{ and } \begin{pmatrix} 2 & -2 \\ -9 & 1 \end{pmatrix},$$

where the matrix entries indicate the number of "points" won or lost for each of the possible outcomes. Payments to *S*s were linked to the number of points won or lost relative to the expected number attainable by ignoring the signal and simply guessing.

At the beginning of each 100-trial block *S* was informed that the signal would be present "less than half the time," "about half the time," or "more than half the time," depending on which relative frequency was to be presented. Immediately following *S*'s response on each trial a number of green or red lights would glow signifying to *S* the number of points won or lost on that trial.

Five *S*s (graduate and undergraduate students) each participated in nine 300-trial sessions following approximately 500 practice trials. One *S* participated in nine additional 300-trial sessions under conditions identical to the first experiment except that signal intensity was increased from .03 to .04 ftl. The relative frequency of signal present was changed after each 100 trials; the payoff matrix was changed after every three sessions. A set of data thus consisted of a total of 300 trials at each of the nine relative frequency-matrix combinations.

The ROC curves specified by Equations 1 and 2 were fitted to each set of data by choosing $\hat{d}$ to minimize

$$Q_1 = \sum_{i=1}^{9} \{\hat{P}_{Si} - 1 + F[F^{-1}(1\text{-}\hat{P}_{Ni}) - \hat{d}]\}^2,$$

and by choosing $\hat{P}_D$ to minimize

$$Q_2 = \sum_{i=1}^{9} \{\hat{P}_{Si} - [\hat{P}_D + (1\text{-}\hat{P}_D)\hat{P}_N]\}^2,$$

where the symbol ∧ over a quantity denotes an empirical estimate.

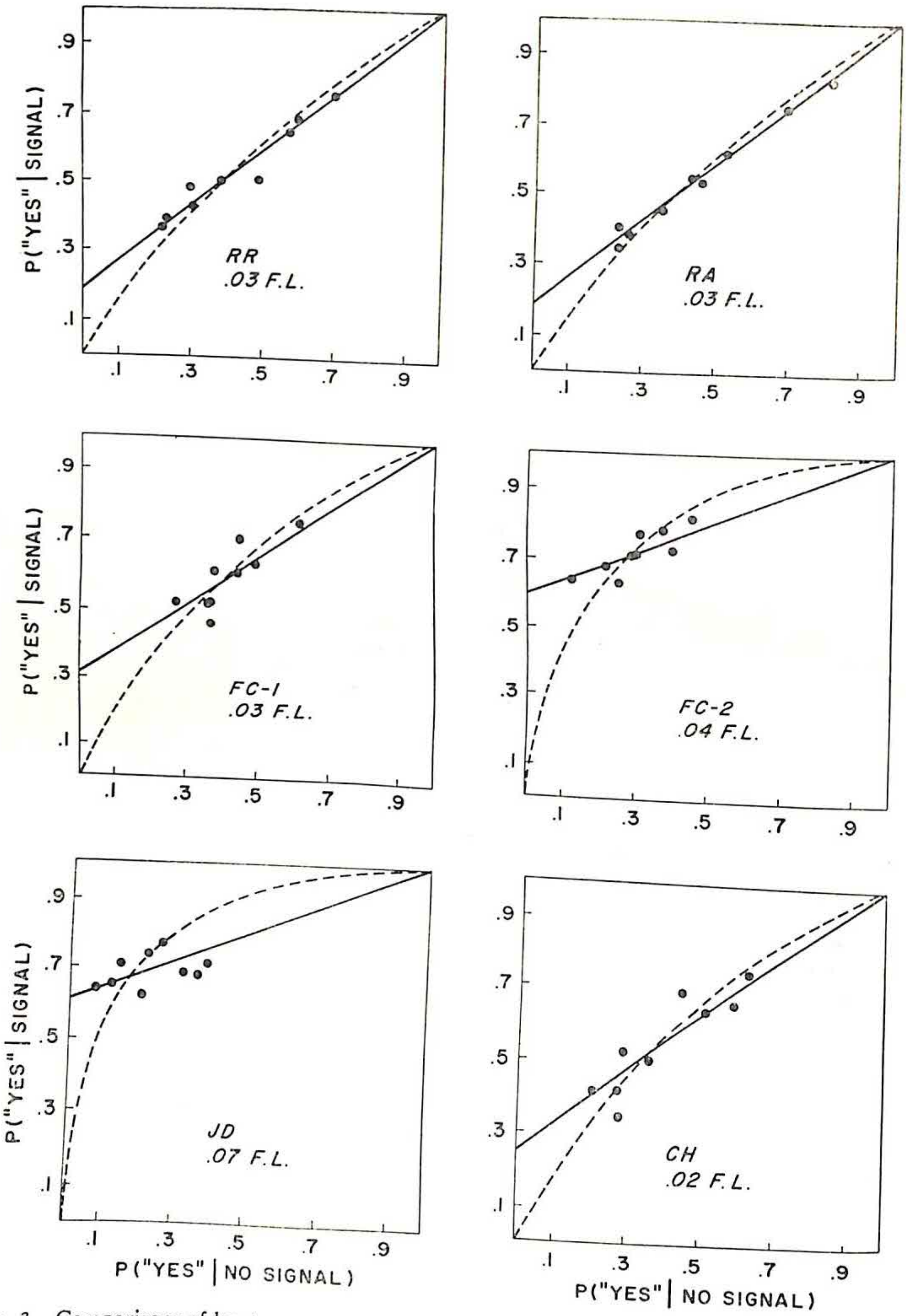


FIG. 3. Comparisons of least-squares ROC curves assuming the simplest form of the theory of signal detection (broken lines) vs. those assuming the Blackwell threshold theory.

As indexes of goodness of fit for the two models, the quantities

$$u = 1 - \frac{Q_j}{\sum_{i=1}^{9} (\hat{P}_{Si} - \hat{P}_S)^2}, \quad j = 1,2,$$

were computed for each set of data. This index, it may be noted, would be a Pearson r^2 if a general linear function were fitted to the data: It is the proportion of variance of $\hat{P}_S$ accounted for by $\hat{P}_N$ with the prediction equation. Although u, like r^2, cannot have a value greater than 1, u can be negative, for the variance of deviations of $\hat{P}_S$ from the fitted ROC curve may be greater than the variance of $\hat{P}_S$ around the mean of the $\hat{P}_{Si}$.

RESULTS AND CONCLUSIONS

It appears both from Fig. 3 and from Table 1 that Equation 2 (the Blackwell theory) provides generally better representations of the obtained ROC curves than Equation 1. These data thus provide no evidence that the theory of signal detection, in the form specified by Equation 1, at least, should generally supersede the relatively simpler Blackwell threshold theory.

The assumptions of Equation 1 may be relaxed—as Swets, Tanner, and Birdsall (1961) have suggested might be necessary for visual stimuli—to permit unequal variances of the noise and noise + signal distributions. In this case, the predicted ROC curve becomes

$$P_S = 1 - F \left\{ \frac{1}{\sigma_{(S)}} [F^{-1}(1 - P_N) - d] \right\}, \quad [3]$$

where $\sigma_{(S)}$ is the *SD* of the noise+signal distribution in units of $\sigma_{(N)}$. As indicated in Table 1, least-squares estimates of the two parameters of Equation 3 from the six sets of data resulted in values of u which are comparable to those for the Blackwell theory. For two *S*s the two-parameter theory of signal detection curves provided closer fits of the data than the one-parameter Blackwell curves. But Equation 3 has the awkward implication that, for high values of P_N, the probability of a "yes" response with no signal is higher than the probability of a positive response with the signal present! Fitting the two parameters to the data from Subject JD, e.g., ($\sigma_S = 6.00$, $d = 3.92$), results in an ROC curve indicating that when $P_N = .95$, P_S is only .82.

TABLE 1

GOODNESS-OF-FIT INDEXES (u) COMPARING THE THEORY OF SIGNAL DETECTION TO THE BLACKWELL THRESHOLD THEORY

S	Blackwell Theory	TSD	
		$\sigma_{(S)} = \sigma_{(N)}$	$\sigma_{(S)} \neq \sigma_{(N)}$
RR	.938	.836	.937
RA	.986	.948	.985
FC-1	.674	.642	.673
FC-2	.657	−.134	.653
JD	−.060	−5.024	.212
CH	.804	.773	.812

REFERENCES

KINCAID, W. M., & HAMILTON, C. E. An experimental study of the nature of forced-choice responses in visual detection. *U. Mich. Engng. Res. Inst. Proj. Mich. Rep.*, 1959, No. 2144–295-T.

SWETS, J. A. *An experimental comparison of two theories of visual detection.* (Doctoral dissertation, University of Michigan) Ann Arbor, Mich.: University Microfilms, 1954, No. 8422.

SWETS, J. A. Detection theory and psychophysics: A review. *Psychometrika*, 1961, 26, 49–63. (a)

SWETS, J. A. Is there a sensory threshold? *Science*, 1961, **134**, 168–177. (b)

SWETS, J. A., & SEWALL, S. T. Invariance of signal detectability over stage of practice and levels of motivation. *J. exp. Psychol.*, 1963, **66**, 120–126.

SWETS, J. A., TANNER, W. P., JR., & BIRDSALL, T. G. Decision processes in perception. *Psychol. Rev.*, 1961, **68**, 301–340.

(Received June 30, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 5, 464–472

STIMULUS FLUCTUATION, REACTIVE INHIBITION, AND TIME BETWEEN TRIALS IN CLASSICAL EYELID CONDITIONING[1]

WILLIAM F. PROKASY[2]

Pennsylvania State University

Predictions from the constructs of reactive inhibition and stimulus-element sampling were tested in human classical eyelid conditioning by manipulating both mean intertrial interval (ITI) and ITI variability. The major findings were that (a) the mean acquisition curve was unaffected by whether or not the ITI was fixed or varied, (b) response probability, latency, and topography were conditional on the duration of the immediately preceding ITIs, and (c) the typical massed-spaced effect appears to result from a scoring bias imposed by a failure to adjust for incomplete latency distributions. Unambiguous support for neither of the tested constructs was obtained, and an interpretation in terms of temporal cues in the situation was suggested.

Two theoretical constructs can account for the fact that, in classical eyelid conditioning, response probability increases more rapidly with spaced than with massed trials. The construct of reactive inhibition (Hull, 1952, p. 9) yields this result because reducing the amount of time between trials also reduces the amount of time available for the dissipation of response-produced inhibition, with the consequence of reduction in overall response probability.

The construct of stimulus-element fluctuation (Estes, 1955a, 1955b, 1959, p. 424) provides a similar prediction, based on the fluctuation of stimulus elements in and out of the population of elements which is available to be sampled on a particular trial. It has been shown (Prokasy, 1961) that this construct, applied as an interaction of elements between the sample selected on a trial and the population from which the sample came, is also adequate to account for the typical massed-spaced effects.

Radically different predictions follow from the two constructs when they are applied to the sequencing of different ITIs. If a set of ITI values, say, 5, 10, 15, and 20 sec., is given for many trials in a random order, it is possible to obtain response probabilities separately for trials following each ITI value. Due to the depressing influence of reactive inhibition there should be an increase in response probability as a function of ITI value. If stimulus-element fluctuation occurs, there will be a correlation, one which decreases with time separating trials, between the samples drawn on successive trials. The result of this initially high, but decreasing, correlation should be a probability decrease as a function of ITI value.

It is unreasonable to expect an experiment to constitute a critical test between the two constructs, simply because they are not mutually exclusive: both, neither, or either

---

[1] This research was supported by a grant from the National Science Foundation, and the paper was written while the author was on a Senior National Science Foundation Fellowship at Indiana University.

[2] I would like to acknowledge the technical assistance of Patricia Creveling, Stanley Lipkin, and George Wolford, all of whom were apprentices in an Undergraduate Research Participation program sponsored by the National Science Foundation.

might account for response variance. If neither is adequate, the results of an experiment cannot tell us this, if only because we must obtain either no significant differences, which rarely educates us, or an outcome consistent in some degree with one or the other. The tactic, then, is to determine which of the two better approximates the results and, wherever possible, attempt to account for the exceptions with the realities of a particular experimental setting.

Results to date have been ambiguous. Prokasy and Whaley (1961) found no correlation between response probability and length of ITI value, while Prokasy and Whaley (1963) found a slight, though reliable, positive relationship. Krauss and Prokasy (1963) found, however, a higher response probability following a 5-sec. ITI than a 50-sec. ITI when just those two values were randomly ordered.

Since the shortest average ITI in the studies to date was 12.5 sec., it may be that a relationship exists but requires a shorter average ITI to be observed reliably. This, then, was the major purpose of the present experiment: to investigate the effects on conditioning performance of manipulating time between successive trials, with emphasis placed on short average ITIs.

METHOD

Apparatus.—The apparatus for picking up, amplifying, and recording eyelid responses is described elsewhere (Prokasy & Whaley, 1961).

The CS was a 450-msec. increase in brightness from 5.2 to 295.6 mL. of a 3-in. milk-glass disk located approximately 32 in. in front of *S*, about 11 in. below eye level. The UCS was a 50-msec. puff of nitrogen delivered to *S*'s right cornea 450 msec. after CS onset. It was delivered through a 4.5-mm. diameter jet adjusted to between .25 and .75 in. from *S*'s eyelash, and was sufficient to support a 175-mm. column of mercury.

Interstimulus interval, CS, and UCS durations were controlled by Tektronix timers. Equipment cues were masked by a 70-db. (re .0002 dyne/cm^2) white noise sound screen which was presented through a speaker mounted in *S*'s cubicle.

Subjects.—The *S*s were 156 volunteers from introductory psychology courses at the Pennsylvania State University who participated in the study to fulfill part of a course requirement. Six were lost through apparatus failure or *E* error.

Experimental design.—The design was a 2 × 3 factorial, with two levels of average ITI and three levels of ITI variability. The average ITIs were 4 and 8 sec. The three levels of ITI variability were zero (i.e., fixed ITI), variable (2, 4, and 6 sec. in the 4-sec. groups and 2, 4, 6, 8, 10, 12, and 14 sec. in the 8-sec. group) and extreme (2 and 6 in the 4-sec. group and 2 and 14 in the 8-sec. group). Group designations will derive directly from the treatment combination: e.g., "4V" will refer to a 4-sec. average ITI with the variable ITI values of 2, 4, and 6 sec., and "8E" will refer to the 8-sec. average ITI with the extreme values of 2 and 14 sec.

The *S*s were assigned, 25 to a group, to the six treatment combinations unsystematically. All were given 295 trials in a single session, with the different ITI values administered an equal number of times (in unpredictable order) within each of the V and E treatments. They were instructed to sit quietly, not to try to figure out the experiment, and to let their reactions take care of themselves. The occurrence of a 1-mm., or greater, pen deflection at least 160 msec. after CS onset, but prior to and overlapping the UCS, defined a CR. Latencies and slopes were measured from the time derivative of the response, which was recorded via an electronic differentiator. The slope index was the maximum displacement (in millimeters) of the derivative pen during the occurrence of a CR, which is, therefore, a measure of the peak speed attained during eyelid closure.

RESULTS

Conventional analyses.—Response probability in blocks of 20 trials (with only 15 trials composing the final block) is plotted in Fig. 1. As anticipated, probability was greater

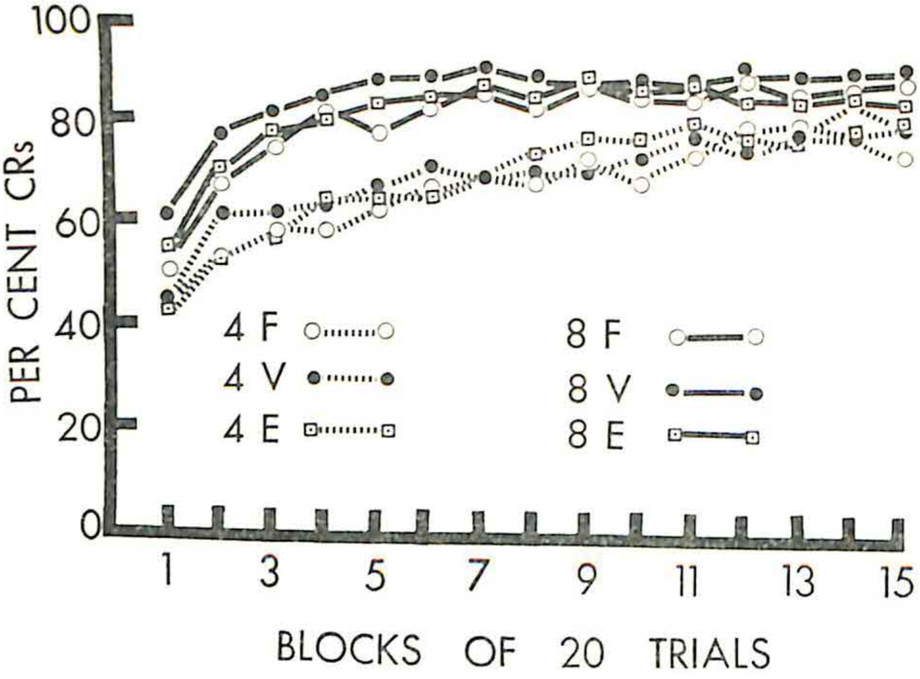


FIG. 1. Response probability across blocks of 20 trials.

for a mean ITI of 8 rather than 4, sec., *F* (1, 144) = 13.12. The performances of the F, V, and E groups, however, are indistinct: Even with a mean ITI as short as 4 sec., whether or not the ITI is fixed has little or no effect on overall response probability, *F* (2, 144) = .27. In addition to the general rise in probability across trials, there was a significant ITI × Blocks interaction, *F* (14, 2016) = 3.98. This reflects the different growth rates of the two average ITIs, but not differences in asymptote: performance in Groups 4F, 4V, and 4E did not stabilize.

Latency, latency *SD*, and slope *SD* were evaluated for each *S* by dividing the trials into blocks of 100, 100, and 95. Latency *SD* and slope *SD* were obtained separately for each *S*, requiring, therefore, a minimum of two responses from each *S* in each block. Two *S*s failed to meet this criterion and cell proportionality was restored by basing subsequent analyses on *N*s of 24 in each treatment combination. The resulting group means and *SD*s are reported in Table 1.

The only significant source of variance with latency as the measure is the interaction between mean ITI and Blocks, *F* (2, 276) = 9.09, which results from a latency decrease over blocks in the 8-sec., but not in the 4-sec., groups.

Latency *SD* was sensitive to both independent variable manipulations. As mean ITI increased, latency *SD* decreased, *F* (1, 138) = 5.94. Latency *SD* was largest in Groups 4F and 8F and smallest in Groups 4V and 8V, *F* (2, 138) = 27.75. Table 1 shows that latency *SD* decreased in the F and V groups as ITI increased, but did not in the E groups, the ITI × FVE interaction being significant, *F* (2, 138) = 5.78. The reduction in latency *SD* across blocks was also significant, *F* (2, 276) = 24.51.

Neither slope nor slope *SD* varied systematically with mean ITI or FVE, although both decreased across blocks, *F* (2, 276) = 16.60 and 13.40

for, respectively, slope and slope *SD*. In addition, there was a Blocks × FVE interaction, F (4, 276) = 3.34, which resulted from an increase in slope *SD*, unique to the V treatment, in the final block.

Effects conditional on ITI.—Since the V groups received either three (4V) or seven (8V) different ITI values, it was possible to assess the average effects of the immediately preceding ITI on momentary response characteristics. For example, the 2-, 4-, and 6-sec. ITI values appeared 98 times each in the schedule for Group 4V. The 295 trials were divided into blocks of 150 and 145, with 50 instances of each interval appearing in the first, and 48 in the last, block. A similar division was made for Group 8V, though there were only 42 instances of each of the seven ITIs over the 295 trials. Table 2 summarizes response probability, latency, latency *SD*, slope, and slope *SD* for the two groups.

Of interest for present purposes the effects of ITI duration.[3] In Group 8V probability increased and latency, latency *SD*, and slope decreased as a function of the duration of the immediately preceding ITI, F (6, 23) = 5.19, 16.13, 3.61, and 4.78, respectively. Although the relationships tended to be less in the second block than in the first, the ITI × Blocks interaction was significant, F (6, 23) = 3.40, only for the latency measure.

Probability increased and latency

---

[3] The statistical results in each subgroup analysis primarily related to block effects are not reported, as they duplicate the results of earlier overall analyses. However, complete summary tables are available from the author.

TABLE 1

MEAN AND *SD* OF LATENCY, LATENCY *SD*, SLOPE, AND SLOPE *SD* OVER THREE BLOCKS OF TRIALS

Group	Latency[a] Blocks			Latency *SD*[a] Blocks			Slope[b] Blocks			Slope *SD*[b] Blocks		
	1	2	3	1	2	3	1	2	3	1	2	3
4 Sec. ITI												
F												
M	313	322	321	69.2	67.9	63.5	9.75	9.55	9.79	3.17	2.71	2.48
SD	51.2	39.7	37.3	13.1	11.1	10.2	2.78	2.57	2.41	1.86	1.26	1.40
V												
M	318	326	324	53.3	47.8	47.1	12.33	11.10	10.66	3.31	2.75	3.50
SD	45.5	43.2	35.4	9.6	7.0	8.1	3.48	2.92	2.74	.67	.67	.71
E												
M	327	324	321	58.9	53.4	52.9	11.57	10.71	10.78	3.53	2.91	2.81
SD	46.0	39.0	37.6	11.0	10.4	11.1	2.76	2.47	2.62	.72	.63	.62
8 Sec. ITI												
F												
M	310	303	299	61.8	54.4	53.3	10.76	9.81	9.90	3.30	2.67	2.85
SD	51.1	47.1	48.1	12.4	12.3	14.3	2.40	2.21	2.19	.71	.62	.57
V												
M	319	309	305	48.7	43.2	45.1	11.82	11.16	10.79	2.99	2.89	2.93
SD	65.2	70.8	59.5	15.4	14.2	8.2	3.23	3.13	2.85	.70	.80	1.44
E												
M	319	316	311	61.6	56.0	55.5	10.59	9.63	9.33	3.16	2.94	2.67
SD	55.4	54.6	53.8	14.9	8.3	9.6	2.60	2.67	2.69	.74	.62	.65

Note.—N = 24 for 4-sec. and 8-sec. ITIs.
[a] Latency and latency *SD* are in milliseconds.
[b] Slope and slope *SD* are in millimeters.

TABLE 2

PROBABILITY, LATENCY, LATENCY *SD*, SLOPE, AND SLOPE *SD* AS A FUNCTION OF THE DURATION OF THE IMMEDIATELY PRECEDING INTERTRIAL INTERVAL IN GROUPS 8V AND 4V

ITI (Sec.)	Probability Block		Latency[a] Block		Latency *SD*[a] Block		Slope[b] Block		Slope *SD*[b] Block	
	1	2	1	2	1	2	1	2	1	2
8V										
2	.749	.845	337	318	50.2	42.8	12.5	11.2	2.79	2.83
4	.806	.852	321	313	48.5	41.3	11.9	10.9	3.01	2.65
6	.815	.872	314	302	44.8	42.6	11.7	11.2	2.87	2.95
8	.841	.882	310	302	42.8	42.0	11.3	10.9	2.66	2.89
10	.849	.872	304	305	41.0	42.4	11.7	10.7	3.01	2.69
12	.838	.888	309	305	44.0	41.6	11.3	10.8	2.79	2.76
14	.818	.878	314	302	37.4	39.1	11.4	10.6	2.77	2.59
4V										
2	.609	.726	328	333	53.2	49.9	11.9	10.7	2.89	2.91
4	.671	.789	317	323	45.8	43.3	12.1	11.1	2.78	3.53
6	.667	.780	318	320	47.8	42.0	12.2	10.7	3.27	3.12

Note.—$N = 24$ for Groups 8V and 4V.
[a] Latency and latency *SD* are in milliseconds.
[b] Slope and slope *SD* are in millimeters.

and latency *SD* decreased as a function of ITI duration in Group 4V, *F* (2, 23) = 5.47, 5.79, and 9.19, respectively. There were no effects on slope (as there were in Group 8V) or slope *SD*.

In Groups 8E and 4E each of the two ITI values appeared 147 times, making it possible to determine the average effects of short runs of each. Thus, for Group 4E the five measures were obtained following exactly one, two, or three successive instances of both the 2- and 6-sec. ITIs. Similar measures were taken in Group 8E following the 2- and 4-sec. ITIs. These data are summarized in Table 3.

In contrast to Group 8V, probability was lower and latency was longer following the 14-sec. ITI than the 2-sec. ITI in Group 8E, *F* (1, 23) = 6.32 and 35.89, respectively. The decreases, as a function of ITI, of latency *SD*, and slope were also significant, *F* (1, 23) = 6.97 and 8.05, respectively. In Group 4E response probability was greater following a 6-sec., rather than a 2-sec., ITI duration, as in Group 4V, but no systematic main effect was observed with the other measures.

Length of run (i.e., number of successive instances of a particular ITI) influenced measures in both Group 8E and Group 4E. In both groups probability increased as run length increased, *F* (2, 23) = 3.59 and 11.93 for Groups 8E and 4E, respectively, though in Group 4E the effect was largely confined to sequences of the 2-sec. ITI and was particularly pronounced in the first block, *F* (2, 23) = 6.40 and 5.09 for the Run × ITI and Run × Block interactions, respectively.

Latency was not uniformly affected by run length in either group. In Group 8E there was a decrease in latency as run length increased for the 2-sec., but not the 14-sec., ITI, which resulted in a significant Run × ITI interaction, *F* (2, 23) = 18.49. The same interaction appeared in Group 4E, *F* (2, 21) = 11.27, but the size of the effect with sequences of the 2-sec. ITI was sufficient to result also in a significant Run main effect, *F* (2, 21) = 19.67.

TABLE 3

PROBABILITY, LATENCY, LATENCY *SD*, SLOPE, AND SLOPE *SD* AS A FUNCTION OF THE NUMBER OF SUCCESSIVE INSTANCES OF ITI VALUES IN GROUPS 8E AND 4E

ITI (Sec.)	Run Length	Probability Block		Latency^{a} Block		Latency *SD*^{a} Block		Slope^{b} Block		Slope *SD*^{b} Block	
		1	2	1	2	1	2	1	2	1	2
8E^{c}											
2	1	.755	.856	321	309	55.8	50.6	11.0	9.7	3.19	2.54
	2	.765	.885	291	290	61.8	50.0	10.4	9.5	2.91	2.65
	3	.804	.867	294	288	58.2	44.4	10.6	9.4	3.20	2.33
14	1	.720	.813	331	328	51.2	46.2	9.9	9.2	2.80	2.75
	2	.754	.810	331	327	51.2	49.2	10.6	9.2	3.02	2.46
	3	.743	.837	334	329	45.2	45.3	10.0	8.7	2.54	2.32
4E^{d}											
2	1	.505	.722	342	333	53.4	54.7	11.5	11.1	3.10	2.88
	2	.640	.774	310	311	58.4	45.2	11.9	11.4	3.31	2.81
	3	.644	.775	316	309	48.7	47.6	11.6	11.1	3.19	2.51
6	1	.657	.835	319	313	54.3	49.3	11.6	11.2	3.06	2.74
	2	.682	.847	320	313	51.6	46.1	11.4	10.9	2.95	2.90
	3	.653	.840	324	308	47.8	49.9	11.4	11.0	3.28	2.73

[a] Latency and latency *SD* are in milliseconds.
[b] Slope and slope *SD* are in millimeters.
[c] $N = 24$.
[d] $N = 22$. $N = 24$ with probability as the measure.

Of the remaining three measures, only latency *SD* provided additional information related to length of run. In Group 8E latency *SD* decreased with run length, $F\ (2,\ 23) = 5.02$, but not uniformly as the Block×ITI ×Run interaction was significant, $F\ (2,\ 23) = 3.82$. In Group 4E, there was a significant Block×Run interaction which reflects a generally decreasing latency *SD* as a function of run length, $F\ (2,\ 21) = 3.66$, with the extent of decrease depending upon the particular block.

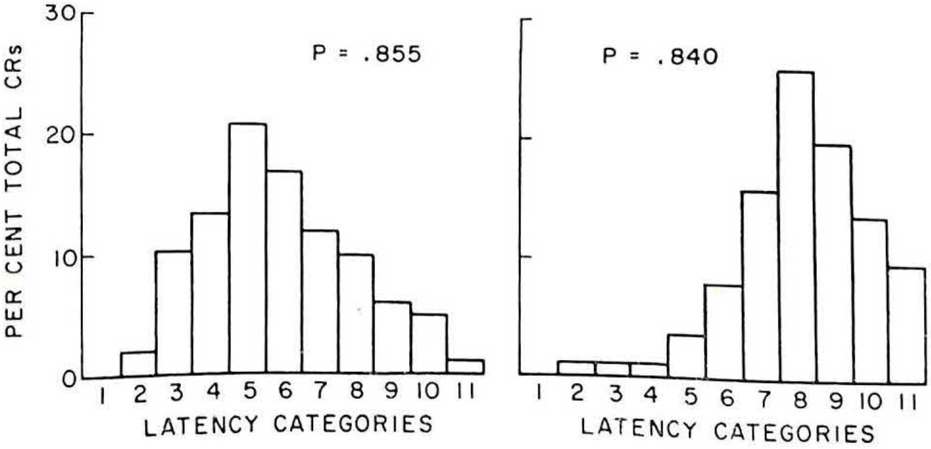


FIG. 2. Two individual *S* latency distributions obtained from Trials 96–295. (There are 11 categories, each 26 msec. in width, beginning with 160 msec. and ending with 447 msec. The *P* values are response probabilities.)

Latency distributions.—Latency distributions for individual *Ss* were obtained over the last 200 trials of the session. Figure 2 shows two of these distributions, both from Group 8V. Approximately 1% of the first *S*'s responses fell in the latency category just preceding UCS onset, whereas approximately 10% of the second *S*'s responses fell in the last latency category. It is apparent that the latency distribution of the second *S* is incomplete and that, therefore, the recorded response probability is biased downward.

The bias illustrated in Fig. 2 is not independent of treatment combination. Mean and median probabilities were obtained for those *Ss* having no more than 5% of their responses falling in the last latency category, and are shown in Table 4. The recording bias is readily apparent with two comparisons. First, significantly fewer *Ss* met the criterion in the 4-sec. groups, χ^2 $(1) = 6.4$, $p < .05$, indicating that more incomplete latency distributions exist in the 4- than the 8-sec. groups. Second, the massed-spaced effect shown in Fig. 1 is not evident in those *Ss* meeting the criterion. In particular, median probability was higher in Group 4E than in Group 8E, Mann-Whitney U $(11, 16) = 21$, $p < .002$.

TABLE 4

RESPONSE PROBABILITY FOR THOSE *S*S WITH NO MORE THAN 5% OF THEIR RESPONSES IN THE FINAL LATENCY CATEGORY

	Mean ITI	
	4 Sec.	8 Sec.
F		
N	11	20
M	.845	.881
Mdn.	.890	.950
V		
N	10	22
M	.970	.939
Mdn.	.983	.935
E		
N	12	17
M	.931	.821
Mdn.	.960	.825

DISCUSSION

Discussion of the overall massed-spaced effect will be deferred, but two other overall results are worth noting. One is that the mean acquisition curve does not reflect whether or not the intertrial interval is fixed or varied. Since response probability is conditional on the duration of the immediately preceding ITI, temporal effects obviously exist. When the intertrial interval is varied, these local effects apparently average out to a level little different from that obtained with a fixed ITI.

The other result is that even though probability rises rapidly, particularly in the 8-sec. groups, there is a steady decrease in the variance of both latency and slope over trials. Furthermore, response slope decreases (i.e., lid closure becomes less rapid) across trials. The conditioned eyelid response, then, shows characteristics not unlike those that accompany the acquisition of a motor skill: an increase in response precision with continued training.

As anticipated, effects conditional on ITI duration were enhanced by adopting much shorter average ITIs. One set of the obtained conditional probabilities follows directly from the inhibition construct: the increase in response probability as a function of the immediately preceding ITI duration in Groups 4V, 8V, and 4E. Other results in the literature which follow from the inhibition construct and which cannot easily be reconciled with the theory of stimulus fluctuation (Estes, 1955a, 1955b) are the increases in response probability from the last trial block of one session to the first block of the next session (Prokasy, Grant, & Myers, 1958; Prokasy & Whaley, 1963) and the slight within-session drops in probability which occur with successive conditioning sessions (Prokasy & Whaley, 1963).

The fine-grain results of Groups 4E and 8E are consistent with the construct of stimulus-element fluctuation, whether between two stimulus populations or between a population and the sample drawn from it: successive instances of the 2-, 6-, and 14-sec. ITIs resulted in an increase in response probability. Furthermore, the increases were greater in the first than in the last trial block, which is also consistent with the construct. These particular predictions result solely from continuous rises in response probability: Response probability should be higher for the second of two successive 2-sec. ITIs simply because more elements have been conditioned prior to the trial following the second than following the first of the two ITIs. Additionally, as probability nears asymptote fewer elements remain not conditioned and, therefore, the effect of differing lengths of ITI should become smaller as, indeed, appears to have been the case in the final trial block.[4] Since response probability following the 2-sec. ITI was greater than that following the 14-sec. ITI in Group 8E, the fluctuation construct provides an adequate summary for all of the probability information available from that group.

If it is assumed that among the stimulus conditions necessary for the occurrence of a CR are included, cues acquired on the basis of the temporal distribution of trials, an alternative interpretation of the data is possible. Such cues would be determined by the discriminability of the particular ITI values employed, the likelihood of particular ITI values succeeding themselves, and the amount of redundancy in the temporal distribution. For example, temporal redundancy is highest with the F (i.e., fixed ITI) treatment and this might well reduce somewhat the effectiveness of CS onset as a determiner of the response.[5] It was in the F treatment that the latency distributions were most heterogeneous. By comparison, temporal redundancy is lowest for the V treatments (and latencies most homogeneous). The amount of time separating the different ITI values is not particularly large, making them relatively indistinguishable, with the net effect that the best information available to *S* is that the likelihood of one trial immediately following another is low but increases rapidly with time. If *S*'s attending behavior paralleled this probability increase the obtained results make sense: greater response probability and shorter and more homogeneous latencies as time between trials increases.

Randomly ordering 2- and 14-sec. ITIs resulted in higher probability and shorter latency following the 2-sec. ITI. Krauss and Prokasy (1963) obtained the same probability effect with random orderings of 5- and 50-sec. ITIs. The characteristic common to both of these distributions is that a trial is either followed immediately (with a probability of .5) by another trial or that a relatively long span of time elapses before another trial occurs. The obtained results are plausible if *S* learns to pay attention immediately following a trial and then, in the absence of another trial, relaxes and is subsequently not ready for one when it does occur.

That latency distributions in human eyelid conditioning are incomplete is not a new finding. Ebel and Prokasy (1963) report individual-*S* distributions which overlap UCS occurrence and other studies (e.g., Gormezano & Moore,

---

[4] The construct of stimulus fluctuation is treated independently of the prediction within statistical learning theory that response probability should asymptote at unity. The value of the construct in no way hinges on that particular prediction, since the assumed fluctuation of elements is unrelated to whether or not they are conditioned.

[5] This was shown more clearly in recent data from our laboratory. Training *S*s with a fixed intertrial interval, and then following just one trial with either a longer or a shorter ITI resulted in sharp drops in response probability on the subsequent trial.

1962) provide group distributions which are curtailed by the UCS. What is new is that whether or not distributions are incomplete depends upon the particular treatment combination. With the usual scoring procedure there was a very powerful massed-spaced effect, consistent with earlier studies (e.g., Prokasy, Grant, & Myers, 1958; Spence & Norris, 1950). But the entire effect was obliterated, or reversed, simply by analyzing the data of *Ss* whose latency distributions were nearly complete. It may be that massed training in human eyelid conditioning does not actually lower response frequency, but only produces changes in the latency distribution. Under these circumstances an arbitrary scoring interval (e.g., 150 msec. after CS onset until UCS onset) can result in wholly different growth rates and asymptotes when nothing but a latency difference exists.

It would be surprising if the effects of an arbitrary scoring technique were confined to the ITI variable. We know, e.g., that UCS intensity (Gormezano & Moore, 1962) and interstimulus interval (Ebel, & Prokasy, 1963; Prokasy, Ebel, & Thompson, 1963) affect latency distributions. In view of present data, it is doubtful that these, or other variables, can be investigated adequately with a scoring method which fails to take into account shifts in location and variance of response latency.

REFERENCES

Ebel, H. C., & Prokasy, W. F. Classical eyelid conditioning as a function of sustained and shifted interstimulus intervals. *J. exp. Psychol.*, 1963, **65**, 52–58.

Estes, W. K. Statistical theory of distributional phenomena in learning. *Psychol. Rev.*, 1955, **62**, 369–377. (a)

Estes, W. K. Statistical theory of spontaneous recovery and regression. *Psychol. Rev.*, 1955, **62**, 145–154. (b)

Estes, W. K. The statistical approach to learning theory. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. McGraw-Hill, 1959. Pp. 380–491.

Gormezano, I., & Moore, J. W. Effects of instructional set and UCS intensity on the latency, percentage, and form of the eyelid response. *J. exp. Psychol.*, 1962, **63**, 487–494.

Hull, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.

Krauss, H. W., & Prokasy, W. F. On intertrial interval discrimination in classical conditioning. *Psychol. Rep.*, 1963, **12**, 138.

Prokasy, W. F. Nonrandom stimulus sampling in statistical learning theory. *Psychol. Rev.*, 1961, **68**, 219–224.

Prokasy, W. F., Ebel, H. C., & Thompson, D. D. Response shaping at long interstimulus intervals in classical eyelid conditioning. *J. exp. Psychol.*, 1963, **66**, 138–141.

Prokasy, W. F., Grant, D. A., & Myers, N. A. Eyelid conditioning as a function of unconditioned stimulus intensity and intertrial interval. *J. exp. Psychol.*, 1958, **55**, 242–246.

Prokasy, W. F., Whaley, F. L. The intertrial interval in classical conditioning. *J. exp. Psychol.*, 1961, **62**, 560–564.

Prokasy, W. F., & Whaley, F. L. Intertrial interval range shift in classical eyelid conditioning. *Psychol. Rep.*, 1963, **12**, 55–58.

Spence, K. W., & Norris, E. B. Eyelid conditioning as a function of the intertrial interval. *J. exp. Psychol.*, 1950, **40**, 716–720.

(Received July 3, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 5, 473-478

PROACTIVE INTERFERENCE AS A FUNCTION OF TIME BETWEEN TESTS

LLOYD R. PETERSON AND ANTOINETTE GENTILE
Indiana University

Short-term retention of individual trigrams was tested in blocks of 6 tests with within-block intervals of 5 or 16 sec. and 91 sec. between blocks. Recall at retention intervals of 0, 3, and 9 sec. was tested. Recall at the 9-sec. interval was a direct function of intertest interval. Intrusions of a letter from the previous trigram at a specific position was inversely related to time between tests. The problem posed for interference theory by a decline in proactive interference over a rest interval was considered.

The rapid forgetting of three unrelated letters (Peterson & Peterson, 1959) or words (Murdock, 1961) after a brief presentation has not been easy to explain in terms of interference. The interpolated activity, counting backward, utilized materials categorically different from the materials to be recalled and presumably should not produce any great amount of retroactive interference (RI). Proactive interference (PI) from previous tests in the session appears to account for some of the forgetting, in that Keppel and Underwood (1962) in their initial experiment using a brief auditory presentation found a significant drop in recall from first to second tests of the session, tests which were used as practice by Peterson and Peterson. However, retention of a third trigram in the Keppel and Underwood study did not differ significantly from the second, and, indeed, recall of the second and third trigrams was no better than that found by Peterson and Peterson in the last quarter of a 48-test session. Retention of the first trigram presented to *S*s was only 50% accurate after an 18-sec. retention interval. Interference from other items in the session could not account for this, and the only other proactive source lay in preexperimental experiences. Relevant to this consideration is the question of what effect the passage of time since learning of previous materials may have on their interfering capacity. If previous learning is forgotten as time passes, then its interfering effects should diminish. This variable has been studied in an immediate memory experiment by Conrad (1960). He varied the time interval between successive tests of digit span over a range of 15-40 sec. No differences in numbers of correct recalls were found, although serial-order intrusions from immediately preceding tests decreased as a function of time. However, recall was tested immediately after conclusion of presentation, and PI can be expected to increase with lengthening of the retention interval (Postman, 1962).

The present experiment investigated the effect of varying the length of the rest interval between tests, using single trigrams presented in a single, brief visual exposure. Instead of following the lead of Keppel and Underwood in their second and third experiments and maximizing PI by increasing exposure time and hence

---

[1] This research was facilitated by Grant GB-704 from the National Science Foundation to Indiana University. Richard Cooley assisted in the investigation.

degree of learning of previous items, PI was maximized by shortening the time permitted for recall. Underwood (1950) has shown that in list learning *PI* is inversely related to the length of time allowed for recall. It was considered that a finding of decreasing PI as time between tests increased would (*a*) be relevant in weighing the plausibility of attributing forgetting in a present experimental session to experiences in the past, and (*b*) offer a problem for interference theory, if it should be found that PI decreased over an unfilled rest interval.

METHOD

The *S*s were 96 students enrolled in introductory psychology classes at Indiana University. They were assigned in order of appearance to a randomized list of experimental conditions. All *S*s received six blocks of six tests each, with 91 sec. elapsing from signal for recall on the sixth test of a block to exposure of the trigram on the first test of the next block. For half the *S*s the intertest interval from signal to recall the trigram of one test to exposure of the next trigram was 5 sec., while for the other half this interval was 16 sec. In either case prior to the start of Block 1 there were nine practice tests with the assigned 5- or 16-sec. intertest interval separating them, and a 91-sec. interval separating the last of the practice tests from the first trigram of Block 1.

The items presented were CVC trigrams of association value below 11% (Archer, 1960). A trigram was shown for slightly less than 1 sec. in a memory drum, followed by exposure of a 3-digit number. The *S*s were instructed to count backward by 3 as fast as possible until recall was signaled by a bell and the going out of the light on the drum paper. The retention interval was either 0, 3, or 9 sec., as timed by a decade interval timer.

For a given *S* each retention interval appeared twice in an irregular order within each block of six tests. Over the first three blocks each retention interval was tested once at each position in the block. Over the group of *S*s equal numbers of *S*s were tested at each retention interval on a particular test. Trigrams were assigned to the six blocks and positions such that within a block a given vowel or consonant was not used more than twice and was separated by at least two tests from its previous use. The order in which the blocks of trigrams were presented was randomized for each *S*.

Only the response given last during the first 4 sec. after signal for recall was scored. In the 5-sec. condition the drum turned to a blank space 4 sec. after the signal for recall, and 1 sec. later turned to the next trigram.

RESULTS

As Fig. 1 shows, the 16-sec. intertest interval resulted in more accurate recall than a 5-sec. intertest interval with either a 3- or 9-sec. retention interval. There was also a decline in recall from first to second position in the block for all save the 0-sec. retention intervals. An analysis of variance was run leaving out the 0-sec. retention intervals because of their extremely low variability. Tests from the first three positions in a block were combined within each condition for each *S*, as were scores from the last three positions. A three-factor analysis resulted, consisting of 3 vs. 9 sec. retention intervals and Positions 1–3 vs. 4–6 as within-*S* comparisons, while the 5 vs. 16 sec. intertest intervals were between-*S* comparisons. Retention Interval was found to be significant, $F\ (1,\ 282) = 355.7$, as were Position, $F\ (1,\ 282) = 8.9$, and Intertest Interval, $F\ (1,\ 94) = 31.0$. The Position × Retention interaction was significant, $F\ (1,\ 282) = 7.3$, as was also the Position × Retention × Intertest Interval interaction, $F\ (1,\ 282) = 4.6$.

The significant interactions reflect the decreasing recall from beginning to end of a block found at the 9-sec. retention interval, which did not appear at the 3-sec. retention interval. Note also that even after a 91-sec. intertest interval, a difference remained between the 5- and 16-sec.

intertest groups at a 9-sec. retention interval which was not found at the 3-sec. retention interval. There is evidently a cumulative effect which did not dissipate as rapidly over the long rest interval for the 5-sec. within-block interval as with the 16-sec. within-block interval, or perhaps there is more to dissipate in the former case.

Evidence for interference from specific items was sought in an error analysis. Intrusions of complete trigrams into recall were infrequent, not exceeding 5% of the tests in any condition. Another type of error occurred more frequently, the erroneous occurrence of a letter in recall of the *n*th trigram which was correct in that same position in the trigram of the $(n-1)$th test. An analysis of these position intrusions, summarized in Fig. 2, shows a low rate of occurrence in the first test of a block with a striking increase in frequency with the next test. Further analysis was carried out by computing for each *S* the conditional probability that a position intrusion occurred given that a correct response had not been made. Positions 2–6 were combined within each condition, and Position 1 excluded because of the paucity of recall failures in that position. Even so nine *S*s in the 16-sec. intertest interval group had to be omitted from the analysis because of a lack of recall failures after a 3-sec. retention interval. Since only one *S* had to be excluded for this reason from the 5-sec. intertest interval group, eight other *S*s from

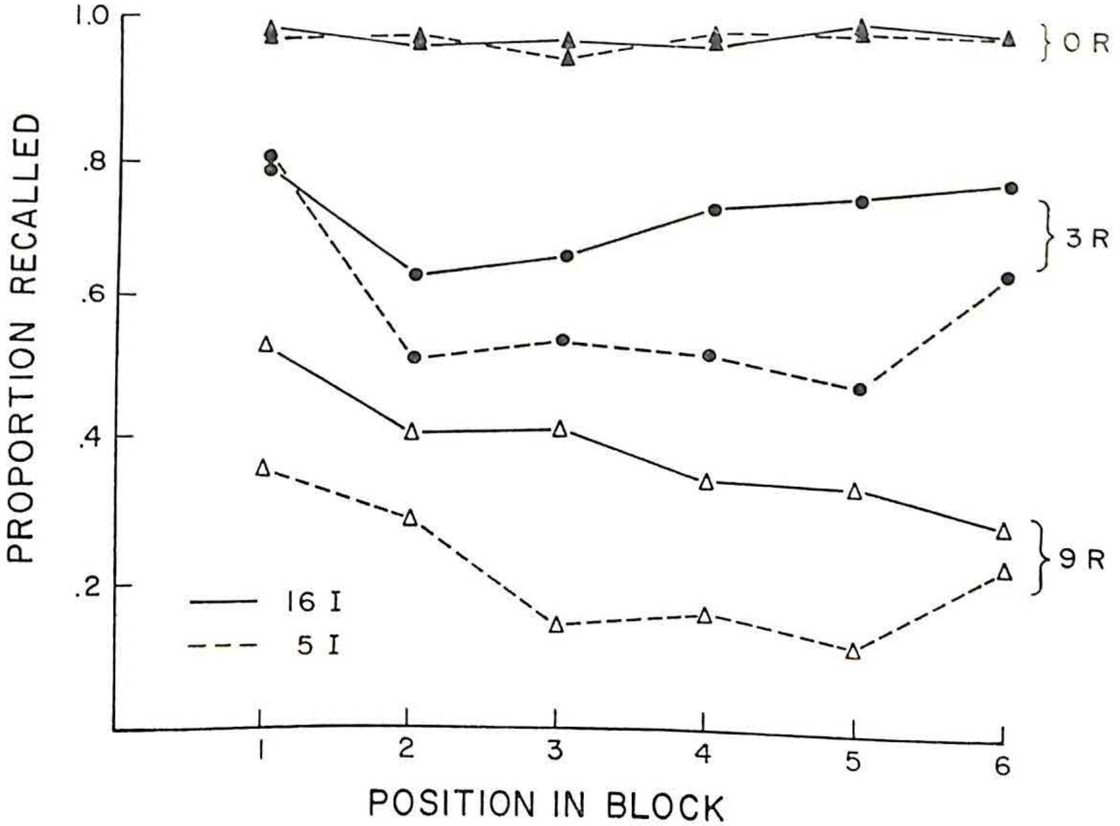


FIG. 1. The proportions of trigrams correctly recalled after retention intervals of 0 sec. (0 R), 3 sec. (3 R), or 9 sec. (9 R) by *S*s having a within-block intertest interval of 16 sec. (16 I) or 5 sec. (5 I).

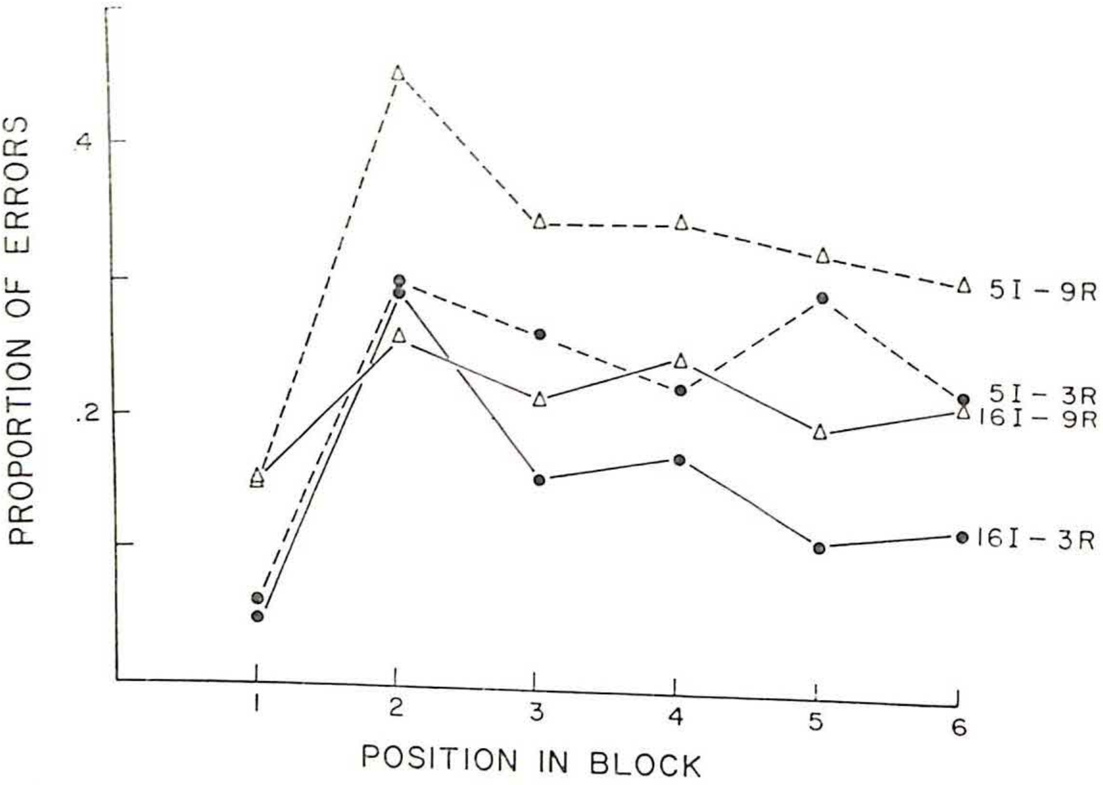


FIG. 2. The proportions of tests in which position intrusions from the previous presentation occurred after a retention interval of 3 sec. (3 R) or 9 sec. (9 R) and within-block intertest intervals of 16 sec. (16 I) or 5 sec. (5 I).

this group who ranked at the top in performance at the 3-sec. retention interval were omitted from the analysis in an attempt to make the two groups approximately equal in ability. The means of the conditional probabilities for the 5-sec. intertest interval group were found to be .55 and .40 for the 3- and 9-sec. retention intervals, respectively. The corresponding means for the 16-sec. intertest interval group were .59 and .35. An analysis of variance found Retention Interval to be significant, F (1, 76) = 23.5. Neither the Intertest Interval nor the interaction was significant. There was found a greater relative frequency of position errors at the 3-sec. retention interval as contrasted to the 9-sec. retention interval in spite of the more frequent recalls found after a 3-sec. retention interval.

DISCUSSION

The number of position intrusions in this experiment decreased over a rest interval. Since these were inversely proportional to the number of correct recalls at corresponding intertest intervals, the interpretation could be made that as the preceding learning is forgotten it produces less interference. A question then arises as to the explanation of the forgetting of the previous learning. It is not immediately apparent that this forgetting is the result of interference. If activity during the rest interval acted as RI to produce this forgetting, then this activity should also have served as PI for the following trigram, and this was not the case.

It is conceivable that the previous trigram was itself subject to attrition from PI with the passage of time. This would explain the decrease in position intrusions over the intertest interval.

However, it would not seem to permit an explanation of the increase in number correct as the length of the rest interval increased. Assuming that the interference consisted of competition at the time of recall, any decline in PI from the previous trigram due to proactive effects on that trigram would be replaced by these other proactive effects from earlier trigrams. Thus, the sum total interference on the current trigram would not be reduced by having the interference come from earlier trigrams instead of the previous trigram. On the basis of this line of reasoning it would not seem plausible that the decline in interference during a rest interval could itself be due to interference.

The relative decrease in position intrusions from the 3- to 9-sec. retention intervals concomitant with a decrease in number correct can be attributed to the counting activity which filled the interval. If interference theory is interpreted to include as interfering activities those not consisting of learning similar in some degree to the learning affected, then the forgetting of previous and current trigrams during the counting interval can be assigned to retroactive interference.

It should be noted that the use of intrusions as evidence of interference has been questioned by Conrad (1960). In an experiment concerned with immediate recall of digits he tabulated serial-order intrusions. These were similar to what have been called position intrusions in the present paper, save that his intrusions were defined in terms of *S*'s response to the previous presentation rather than to the presentation itself. Conrad found that serial-order intrusions decreased with increase in intertest interval even though the number of correct recalls remained at the same level. There would not seem to be any serious discrepancy between Conrad's results and those of the present study when it is considered that he tested at a short retention interval which when averaged over all digits in a sequence would probably be closer to the 3-sec. condition of the present experiment than the 9-sec. condition. Conrad suggests that failure of recall occurred independently of interference from previous tests, and the previous digit was substituted to fill the gap. Applying Conrad's view to the present experiment, the increase in number correct with increasing rest intervals could be attributed to performance factors, while the decline in position errors over the same intervals was a matter of decay in time. While this may be true in part, it does not seem reasonable to hold that occurrence of a previous trigram or part of it may not have some deleterious effect on recall of the present trigram. Thus, we suggest that some forgetting of the previous trigram does occur over a rest interval, but when such a trigram does occur it interferes with recall of the correct response. The nature of the events filling an interval undoubtedly determine the rate of forgetting, more similar activity producing more rapid forgetting. But, any interval not filled with rehearsal appears to produce forgetting to some degree. Therefore, the interpretation of Keppel and Underwood's first experiment, in which significant forgetting was found on the first test of a session, seems to involve the effect of the interval of time spent in counting.

These conclusions must be limited to the case of retention after a single brief presentation, for they could not be applied without further experimentation to presentations of longer duration, as in the second and third experiments of Keppel and Underwood. In the latter case retention was close to 100% on the first test of a session, suggesting that the counting had little effect. In this case a rest interval would not seem likely to have any great effect on retention either, and PI might be expected to extend over longer intervals. However, the greatest retention loss in their second experiment was from the first to the second test, and the added loss on the third test was small. The third study of Keppel and Underwood, which was interpreted as showing continued large decrements in recall through six tests, suffers from a methodological defect. A given *S* was tested at retention intervals of 3 and 18 sec. in strict alternation. After the first test, a 3-sec. interval was always followed by an 18-sec. interval, and vice versa. Given a decreasing retention function on the second test due to PI, the difference between the 3- and 18-sec. retention conditions would be expected to increase progressively as a result of the greater frequency of successful recalls after a 3-sec. interval in contrast to an 18-sec. interval. Assuming that a correct recall increases the likelihood of later occurrence of the response, as well as setting ahead the zero point on the forgetting curve, greater interference from the previous trigram should result after a 3-sec. retention interval than after an 18-sec. retention interval. No valid conclusion can, therefore, be drawn from that experiment concerning the course of PI.

REFERENCES

ARCHER, F. J. Re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychol. Monogr.*, 1960, **74**(10, Whole No. 497).

CONRAD, R. Serial order intrusions in immediate memory. *Brit. J. Psychol.*, 1960, **51**, 45–48.

KEPPEL, G. K., & UNDERWOOD, B. J. Proactive inhibition in short-term retention of single items. *J. verbal Learn. verbal Behav.*, 1962, **1**, 153–161.

MURDOCK, B. B. The retention of individual items. *J. exp. Psychol.*, 1961, **62**, 618–625.

PETERSON, L. R., & PETERSON, M. J. Short-term retention of individual verbal items. *J. exp. Psychol.*, 1959, **58**, 193–198.

POSTMAN, L. The temporal course of proactive inhibition for serial lists. *J. exp. Psychol.*, 1962, **63**, 361–369.

UNDERWOOD, B. J. Proactive inhibition with increased recall-time. *Amer. J. Psychol.*, 1950, **63**, 594–599.

(Received July 6, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 5, 479–489

SHORT-TERM RECOGNITION MEMORY FOR SINGLE DIGITS AND PAIRS OF DIGITS [1]

DONALD A. NORMAN AND WAYNE A. WICKELGREN
Harvard University *Massachusetts Institute of Technology*

The operating characteristic is used to examine the relation between the recognition of a single item and the recognition of a pair of items. 29 Ss listened to a sequence of 5 digits, then copied a sequence of 8 digits, and then were given a test of recognition memory for 1 or 2 digits from the original sequence. The operating characteristic for single digits is a smooth function that is symmetrical about the major diagonal, whereas the curve for pairs is highly asymmetrical. False-recognition rates for test pairs containing 1 digit from the original sequence are only slightly greater than false-recognition rates for completely new pairs. Recognition of a pair does not appear to result from independent recognition of each digit. A mathematical model is developed in which the strength of the memory trace has a continuous distribution which is incremented in a probabilistic fashion upon presentation of an item or pair.

A recognition experiment is a choice experiment, for which powerful analytical tools exist (Bush, Galanter, & Luce, 1963; Swets, Tanner, & Birdsall, 1961). The analytical tools developed in psychophysics can be applied to the results of many substantive areas so long as the data come from experiments that meet certain formal requirements. Experiments in recognition memory, as opposed to experiments in recall, meet these requirements.

In an experiment in recognition memory *S* is asked to judge whether a test item appeared in the original list (and is an *old* item) or whether it is *new*. In a choice experiment such as this, changes in *S*'s response biases can have as much effect on his performance as changes in the stimulus conditions. A conservatively biased *S* might respond "yes" only if he were certain that the test item was old, whereas a liberally biased *S* might always respond yes unless certain that the test item was new.

We consider here a rather general two-stage model of the recognition process in which two systems operate to determine the response: a memory and a decision system. Stimulus items are represented in the memory system by any one of a number of forms—the actual process need not concern us at this time. When the recognition test takes place, the decision system must select a response using, as a basis, the information provided by the memory system. The output of the memory system can be characterized as a reflection of the strength of the test item in the storage system. (At this time it is important to note that even a new test item may have some strength in the memory system.) The decision function maps these strengths onto responses by comparing the strength of the test item with some criterion value. Items that exceed the criterion receive a response of yes, otherwise they are

---

[1] This work was performed while Donald Norman was a National Science Foundation Postdoctoral Fellow at the Harvard Center for Cognitive Studies and was also supported by a grant to Wayne Wickelgren, MH 08890-01, from the National Institute of Mental Health, Public Health Service.

assumed to be unfamiliar (new) items and receive a no response. In this scheme, then, false recognitions contain valuable information. By forcing *S* to vary his biases (and thereby his criterion strength) while holding the stimulus conditions constant (and, hopefully, holding constant his memory of the items) we trace out the relative strengths of the distributions of old and new items. The relation between the relative amount of hits and false recognitions in psychophysics is called the receiver-operating characteristic (ROC curve). To emphasize the study of memory rather than signal detection, we call our curves the memory-operating characteristic (MOC curve).

In the binary-choice experiment just described each point must be obtained in what is essentially a separate experiment in which *S*'s biases are manipulated by changing his instructions or payoffs. Fortunately, however, a more economical technique has been developed (Egan, Schulman, & Greenberg, 1959; Pollack & Decker, 1958). After making his binary decision an *S* can indicate his confidence in that decision on a rating scale. We interpret his confidence as a direct reflection of the strength of the item along a unidimensional scale from a "most confident no" to a "most confident yes." This permits us to get several points of the operating characteristic in one experiment.

The present study applies the operating characteristic to examine the relation between the recognition of a single item and the recognition of a pair of items. In particular, we are concerned with the following problems:

1. Is the strength of the memory trace two valued, multivalued, or continuous? Is the number of values different for single items and pairs of items?

2. How does the presentation of an item or pair of items transform its strength in memory? That is to say, what is the relationship between the old and new distributions for single items and for pairs of items? Is the relation different for single items and pairs?

3. How does the nature of the test pair affect the recognition of pairs? In the case of single items there are only two types of test items: new and old. In the case of pairs there are many more possibilities: Test pairs might have both items old and in the original order, both items might be old but in the reversed order, test pairs might have only one item old, or neither item might be old. Does the false-recognition rate for pairs increase with increasing similarity of the test pair to a presented pair? If it does, then it would appear that recognition of a pair results from independent recognition of its components. If it does not, then perhaps what is recognized is the association between the items or a concept associated strongly to the pair, but weakly to each item.

METHOD

Procedure.—The procedure for each trial was as follows: *Presentation*: A random sequence of five different digits was presented at the rate of ½ sec/digit. The *S*s listened to these digits and attempted to remember them. *Interference*: A random sequence of eight different digits (though not all different from the first set of five digits) was presented at the rate of ½ sec/di it. The *S*s copied these digits as they were presented and covered with a blank card the digits they copied. *Probe*: A single digit or a pair of digits was presented. The *S*s decided whether this digit or pair of digits was in the original sequence of five digits and indicated their confidence in this decision on a 5-point rating scale. The *S*s were given 15 sec. on each trial to indicate their decision and confidence. A "ready" signal was given at the beginning of each trial, and a buzzer was sounded during the 1-sec. intervals between presentation and interference and between interference and probe. A trial required about 25 sec. altogether. The entire experiment was recorded on tape.

When the probe was a single digit, *S*'s decision was yes or no. When the probe was a pair of digits (e.g., 85), *S*s chose one of the following four alternatives: (85) both digits were in the original sequence in the same order as in the probe, (58) both digits were in the original sequence but in the opposite order, (5 or 8) only the one indicated digit was in the original sequence, (no) neither probe digit was in the original sequence. The *S*s were assured that if both digits were "old" (i.e., in the original sequence), then they had been immediately adjacent in the original sequence in one order or the other.

Design.—There were 20 conditions defined by the relation of the probe to the presented sequence. In 5 conditions the probe was a single digit that had occurred in the presented sequence, 1 condition for each serial position in the presented sequence. Since the probe consisted of one "old" digit, we refer to this as Cond. *o*. Condition *n* refers to cases where the probe was a single "new" digit (i.e., the digit did not occur in the original sequence of five digits). Condition *n* occurred as often as Cond. *o*.

In four conditions the probe was a pair of digits that had occurred in the presented sequence as a pair of adjacent digits in the same order as the probe; there is one condition for each serial position for a pair. Since the probe consisted of two old digits in the same order, we refer to this as Cond. *oos*. In four conditions the probe was a pair of old adjacent digits but in the reverse order from their order in the presented sequence: this is Cond. *oor*. In five conditions the probe was a pair of digits, one old and one new: this is Cond. *on*. In one condition the probe was a pair of new digits (Cond. *nn*); this condition occurred as often as Cond. *on*.

There were 18 conditions replicated once each per block of 28 trials and two control conditions (*n* and *nn*) replicated five times each per block. Thus, there were 28 trials per block, and three blocks in the experiment, for a total of 84 trials. Conditions were ordered randomly in each block. All three blocks were different. The *S*s were 29 M.I.T. undergraduates who participated in the experiment to fulfill a requirement of their psychology courses. The 29 *S*s were run in two groups of about equal size. Total time for the experiment was about 50 min.

Data analysis.—When the probe is a single digit (o or n) S must choose 1 of 10 decision-confidence pairs. Let $i = 1, 2, \cdots, 5, 6, \cdots, 10$ stand for the responses running from yes with confidence "5" (greatest confidence), through yes with confidence "1" (least confidence), no with confidence "1," and finally to no with confidence "5." Let $f_i(x)$ represent the total frequency (over all three blocks, all serial positions for that condition, and all 29 *S*s) with which response i occurred in Cond. x. Let $r_i(x) = f_i(x) / \sum\limits_i f_i(x)$ represent the relative frequency with which response i occurred in Cond. x. Let $R_i(x) = \sum\limits_{j=1}^{i} r_j(x)$ represent the cumulative relative frequency with which responses 1 through i occurred in Cond. x.

Using the above definitions it is easy to compute two functions that will be used in the analysis of the data, the MOC curve and the a posteriori probability of a response. The MOC curve is a plot of $R_i(x)$, the "hit rate" for some Experimental Cond. x, against $R_i(y)$, the "false-recognition" rate for some Control Cond. y. Since the MOC curve is a plot of two cumulative-probability functions, one against the other, the curve originates at (0, 0) and ends at (1, 1). The shape and the area under the curve are the principal properties of the MOC curve.

The ratio, $r_i(o)/[r_i(o) + r_i(n)]$, is the a posteriori probability that the probe item was old when the response of S was i. If the memory trace has many different degrees of strength over some range of values of i, then the a posteriori probability will be a monotonic decreasing function of i, for values of i in this range. If the memory trace has a more or less constant strength for several values of i, then the a posteriori probability for these values will be constant.

The analysis of single-digit conditions is easier than the analysis of pair conditions because with pairs there are four, instead of two, possible decisions. However, for our purposes it is possible to classify these four decisions into two categories (yes or no). Having done this, the data are in the same form as the data for single digits, and we can use the previously described analytic functions for both pairs and single digits. There are two meaningful ways of classifying these four decisions into two decisions, ordered-pair analysis and unordered-pair anlaysis. To analyze recognition memory of *ordered pairs* we count a decision as yes only if *S* responded that both digits were old and in the *same* order as the probe. This is called "*ordinary scoring*." In "*reverse scoring*," a decision is counted as a yes response only if *S reversed* the order of the probe. To analyze recognition memory of *unordered pairs*, we count a decision as yes if *S* responded that both probe digits were old, regardless of the order of the digits in his response.

Theory.—Aside from the purely empirical question of whether pairs are remembered differently from single digits, there are certain theories of the nature of the memory trace that can be evaluated using the data of the present experiment. One theoretical issue with which we are concerned in this paper is whether the memory trace is two valued, multivalued, or continuous. With a continuity model the memory trace may take on all degrees of strength over some range. If we assume that the memory traces for both old and new items have overlapping normal distributions, then we have a "normal-continuity model" formally identical to Thurstone's Law of Comparative Judgments (1927) and the decision aspects of signal-detectability theory in psychophysics. The normal-continuity model predicts MOC curves that are smooth functions with continually changing slope, which when plotted on normal-normal probability paper become straight lines. This model also predicts that the a posteriori probability function will be monotonic decreasing. Similar predictions would be made by any model which assumed that the strength distributions for old and new items were continuous and unimodal. Therefore, one could refer to this model as the "unimodal-continuity model."

The two-valued strength model of memory is formally equivalent to the two-state threshold model in psychophysics (Luce, 1963). The two strengths are "full strength" (above threshold) and "no strength" (below threshold). In the threshold model it is possible for some new items to have full strength. This model predicts MOC curves that consist of two intersecting straight-line segments ending at (0, 0) and (1, 1), where the probability that an old item exceeds the threshold (p) and the probability that a new item exceeds the threshold (q) defines the intersection of the lines at coordinates (q, p). If there are but two states for the memory trace, then *S*s cannot make the fine distinctions required of them by the category judgments, and the a posteriori probability function will consist of two horizontal straight-line segments, one above the threshold having the value $p/(p + q)$ and one below the threshold having the value $(1 - p)/[(1 - p) + (1 - q)]$ (Nachmias & Steinman, 1963).

RESULTS

Comparison of single digits and pairs.—The MOC curve for single digits is shown in Fig. 1A and 1B. The ordinate is the hit rate estimated by the average of all five serial positions of Cond. *o* [$R_i(o)$]; the abscissa is the false-recognition rate estimated from Cond. *n* [$R_i(n)$]. The data are shown on linear coordinates in Fig. 1A; Fig. 1B contains the same data plotted on normal-normal probability coordinates. Figure 1 also shows the two MOC curves of Cond. *oos* averaged over all four serial positions for pairs [$R_i(oos)$], vs. the Control Cond. *nn* [$R_i(nn)$], analyzed as both ordered and unordered pairs. In Fig. 1A the smooth curve is the best-fitting prediction of the normal-continuity model for the single digits; the two straight-line functions are the best-fitting predictions of the two-valued strength model to the ordered- and unordered-pair data. The normal-continuity curves were obtained by visually estimating the best-fitting straight lines to the data on normal-normal coordinates (Fig. 1B) and transferring the estimated function to Fig. 1A. The curves for the two-valued strength model were obtained by visually estimating the best-fitting pair of straight lines to the data of Fig. 1A, with the restriction that the lines pass through the coordinates (0, 0) and (1, 1).

The curve for single digits appears to be a smooth function of continually changing slope that is symmetrical about the major diagonal. The fact that the MOC curve for single digits is a straight line with a slope of approximately 1 in Fig. 1B implies that these results could have been

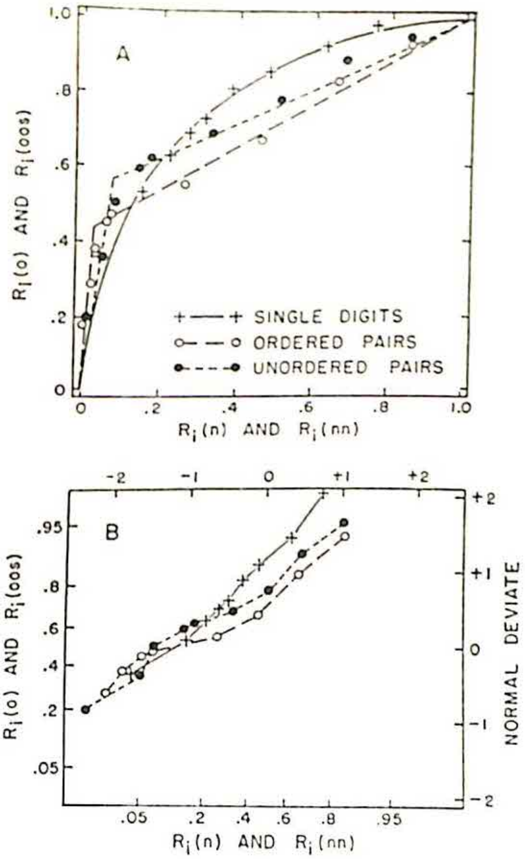


FIG. 1. MOC curves for single digits, ordered pairs, and unordered pairs. (The data are plotted on regular, linear coordinates in A and on normal-normal probability coordinates in B. In A the smooth curve is the best-fitting prediction of the normal-continuity theory to the data for single digits, and the straight-line functions are best-fitting predictions of the two-valued strength model to the data for pairs.)

obtained if the strengths of old and new items had overlapping normal distributions with equal variances.

The curves for pairs are highly asymmetrical and can be closely approximated by the two intersecting straight lines of the two-valued strength model. The data are fit very poorly by normal-continuity curves, even without the restriction that the variances of two underlying distributions be equal. It is difficult to assess visually the fit of the straight lines to the data, particularly above the threshold (the "lower limb" of the function) where the correct-recognition rate is changing very rapidly relative to the false-recognition rate. The a posteriori probability function provides a more sensitive test for the fit of the two-valued strength model.

The a posteriori probabilities for ordered and unordered pairs are shown in Fig. 2A and 2B, respectively. The straight horizontal lines are the theoretical predictions of the two-valued model; and it is clear that, above the threshold at least, they do not describe the data. The data do suggest, however, that although the response categories above threshold do convey information, those below the threshold do not. Thus, even though we are forced to discard the two-valued threshold model, we cannot discard the idea of a threshold.

MOC curves for ordered and unordered pairs, analogous to those of Fig. 1A, can be obtained for Cond.

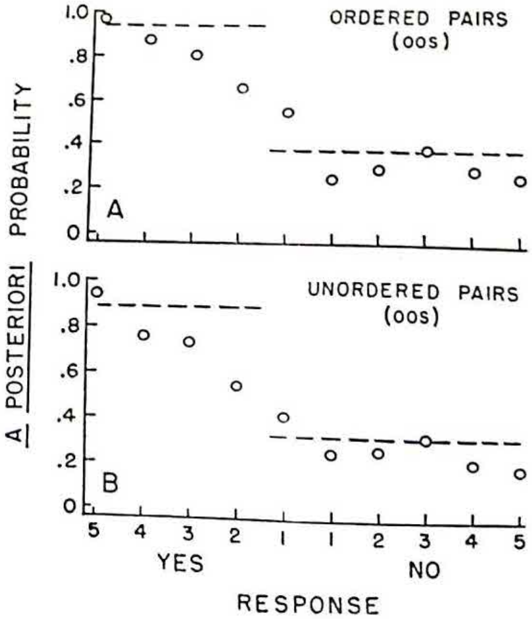


FIG. 2. A posteriori probability functions for ordered and unordered pairs. (The dotted lines are the theoretical predictions of the two-valued strength model.)

oor. The method of analyzing the unordered-pair data of Cond. *oor* is identical to the method of analyzing the unordered-pair data of Cond. *oos*. In the ordered-pair analysis for *oor*, however, a response is scored as yes, if and only if *S* correctly reverses the order of the probe (reverse scoring). Figure 3A compares the MOC curves for the ordered-pair analysis of Cond. *oos* and *oor*. MOC curves for the unordered-pair analysis of Cond. *oos* and *oor* are compared in Fig. 3B. The curves for the two conditions are very similar in shape and location, allowing us to draw the same conclusions about the underlying processes of recognition memory in the two conditions.

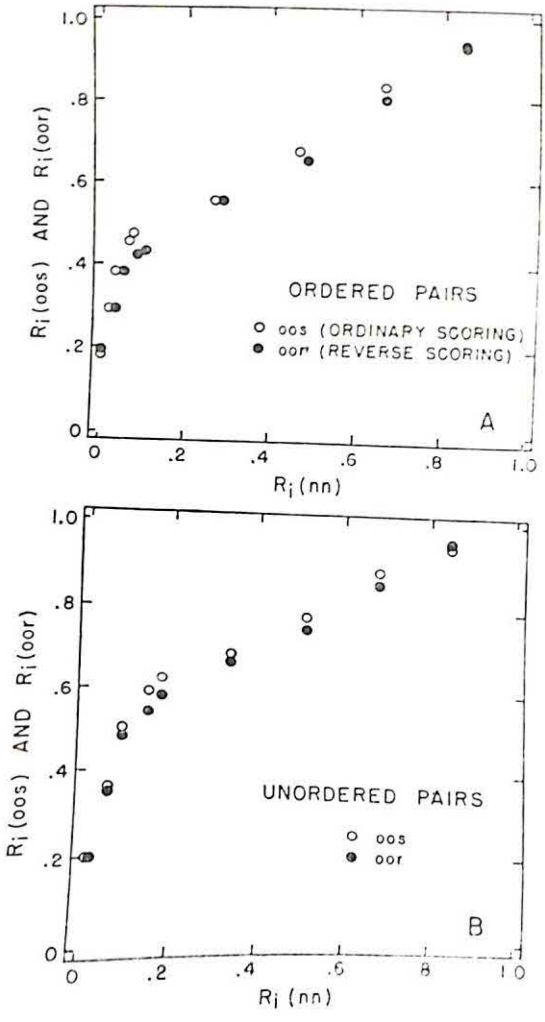


FIG. 3. MOC curves for Cond. *oos* and *oor*.

Correct recognition of pairs, either ordered or unordered, is superior in Cond. *oos*, but the difference between the functions for *oos* and *oor* is very small. It is rather surprising that it makes so little difference for the recognition of pairs whether the probe is in the same or reversed order as the initial presentation.

Degree of oldness and false-recognition rates.—In the preceding section the false-recognition rate for pairs was determined for Cond. *nn* using *ordinary* scoring. Using ordinary scoring in the analysis of ordered-pair recognition, false-recognition rates can also be determined for Cond. *on* and *oor*. Using *reverse* scoring in the analysis of ordered-pair recognition, false-recognition rates can be determined for *on* and *oos*. In the analysis of unordered pairs, the false-recognition rates can also be determined for Cond. *on*.

Using ordinary scoring and ordered-pair analysis, the pair conditions were ranked as follows according to increasing "degree of oldness": *nn*, *on*, *oor*, and *oos*. In Cond. *nn*, neither digit is old; in *on* one of the digits is old; in *oor* both digits are old but in a new order; in *oos* both digits are old and in the old order. Using reverse scoring and ordered-pair analysis the pair conditions were ranked in increasing degree of oldness in the following way: *nn*, *on*, *oos*, and *oor*. The ranking of degree of oldness depends on the oldness of the response that is scored as a yes, not on the oldness of the probe pair. With reverse scoring the inversion between *oor* and *oos* occurs because the probe pair must be reversed to qualify as a yes response.

The lower three curves in Fig. 4A show the cumulative false-recognition rates for each response category using ordinary scoring and ordered-pair

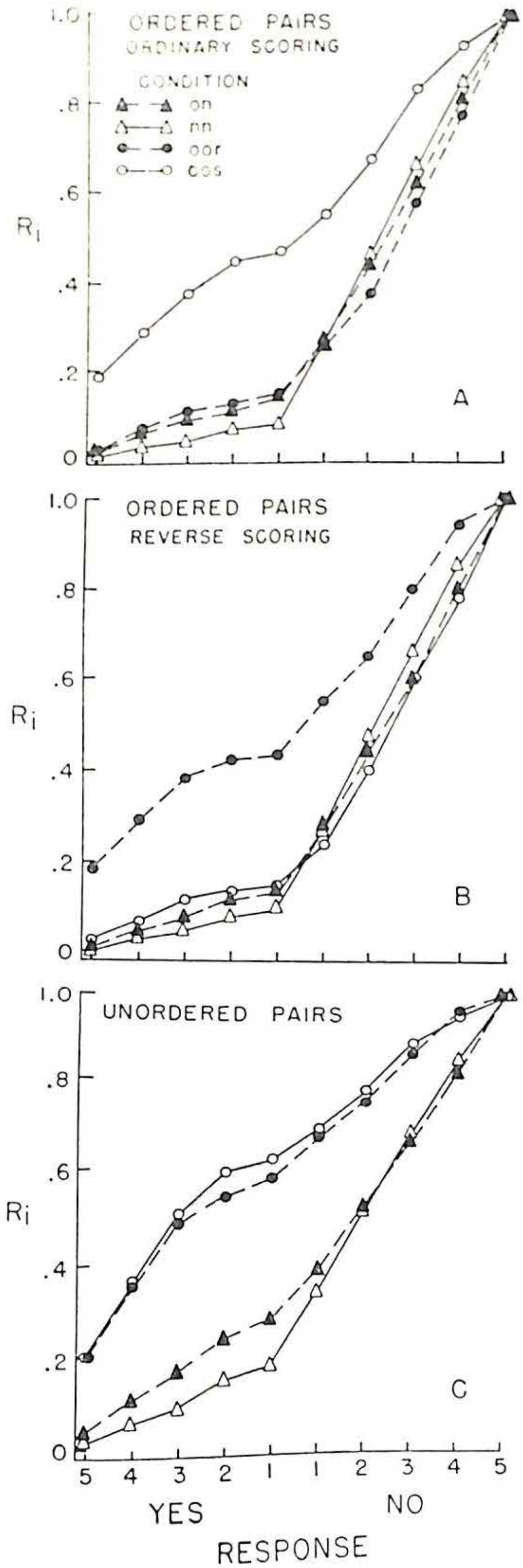


FIG. 4. Recognition rates for different degrees of oldness of the test pair.

analysis. The false-recognition rates are determined for conditions where the yes responses have three different degrees of oldness—namely, *oor*, *on*, and *nn*. If *Ss* used partial cues to help in the recognition task we might expect the results to be ordered by degree of oldness: the greater the degree of oldness, the more false recognitions. The results of the ordered-pair analysis show that this effect does indeed occur, although the magnitude of the effect is very small. On the bottom portion of the curves (where *Ss* are responding yes with various degrees of confidence), the greater the degree of oldness, the higher the curve. On the upper portions of the curves (where *Ss* are responding no with various degrees of confidence) the situation is reversed: the higher the degree of oldness, the lower the curves. These results indicate the degree of oldness does not help much in deciding whether to say yes or no, but *Ss* generally are more confident of either a yes or a no decision as the degree of oldness increases.

Figures 4B and 4C present data analogous to that in Fig. 4A. In Fig. 4B we used reverse scoring for ordered-pair analysis. In Fig. 4C we used unordered-pair analysis. The relationship between degree of oldness and false-recognition rate in Fig. 4B and 4C is the same as that in Fig. 4A. No statistical test has been developed to determine if two operating characteristics are significantly different. However, for the curves in Fig. 4A, 4B, and 4C it is possible to use the Kolmogorov-Smirnov two-sample test to determine if the maximum difference between any two cumulative-probability functions is significant. In this case a maximum difference of .10 between two false-recognition curves will be significant at the .05

level. The maximum differences in Fig. 4A, 4B, and 4C are .09, .07, and .10, respectively. Only the latter difference is significant at the .05 level, though the other differences are close to significance.

The upper curve in Fig. 4A and 4B and the upper two curves in Fig. 4C are the correct-recognition rates for ordered-ordinary, ordered-reverse, and unordered analyses, respectively. It is clear that the differences between the correct-recognition curves and any of the false-recognition curves are far greater than any of the differences among the various false-recognition curves. The fact that we did not obtain a gradual change in recognition rate with degree of oldness indicates that *S*s do not remember a pair of digits by the first digit, the second digit, and the order independently. Memory for a pair is clearly not an additive function of the memory for its parts.

Practice and serial-position effects.—To examine the effects of practice, we computed separate MOC curves for each of the three blocks of the experiment. The results show a definite improvement in recognition memory for single digits over the course of the experiment. However, the opposite practice effect is obtained in recognition memory of ordered pairs. There is no obvious interpretation of this difference. We can only suggest that the practice effects may result from the differential operation of proactive inhibition or a change in strategy of *S*s during the experiment. The reliability of the average MOC curve for a condition is underscored by the fact that the MOC curves for a given condition in each of the three blocks of the experiment are similar in shape and differ only in height.

The conventional serial-position effect was obtained for both pairs and single digits, but the effect was much more pronounced for pairs.

DISCUSSION

The findings for recognition memory of pairs of digits are rather different from the corresponding findings for recognition memory of single digits. There is a bigger serial-position effect for pairs than for single digits, the practice effects are opposite in the two cases, and the MOC curves have very different shapes. Logically, the process of recognizing ordered pairs must involve more than the single-digit recognition process applied to each of the two digits, viz., it must involve the recognition of order. However, the recognition of unordered pairs could result from the independent recognition of each digit of the pair. It is possible to test this hypothesis using the MOC curves for single digits and unordered pairs by making some assumptions about the way an *S* combines the recognition of single digits into the recognition of pairs. One reasonable assumption is that *S* assigns to the pair the decision-confidence rating that reflects the strength of the weakest of the two digits. This model is wrong. It predicts an MOC curve for pairs that is similar in shape to the curve for single digits and substantially above it. This means that the predicted curve has a shape and location entirely different from the empirically obtained curve for pairs. Moreover, several other models of the way that *S* might combine single-digit recognition into pair recognition fail in similar ways. The conclusion that pair recognition is not based on single-digit recognition is further supported by the fact that degree of oldness of a pair has so little effect on the false-recognition rate.

Granted that pair recognition is a different process from single-digit recognition, what relation obtains between the recognition of a pair and the single digits that make up the pair? Do pairs containing common digits (94, 49, 91) differ from each other in the same way as pairs with no digits in common (94, 83)? The present findings indicate that they do, almost. The false-recognition rates for two old items in reverse order, one old item and one new item, and two new items are almost identical, although the ordering of these conditions is consistent in the direction of higher false-recognition rate for higher degrees of oldness. The consistency of the ordering should not detract from the more important fact that the differences in false-recognition rate are extremely small.

It has already been shown that the normal-continuity model is consistent with the MOC curves for single digits, although other continuous distribution functions for old and new digits might also describe these data. The important result is that the memory trace for a single digit may have one or many different degrees of strength rather than only two or three. It is completely consistent with the present findings to assume that the distribution of single-digit trace strengths is a continuous function of a real variable. It must be noted, however, that the present findings are obtained by averaging over all *S*s. If there are substantial individual differences in the trace distributions and the characteristics of the decision systems, then it is quite possible to obtain smooth average MOC curves from individual MOC curves of very different shape and consistent with other models.

There is less need for caution in interpreting the MOC curve for pairs because such a curve is highly unlikely to result from averaging individual curves of very different shape. The straight-line portion of the MOC curve for pairs is inconsistent with the normal-continuity model, no matter what one assumes about individual differences. However, as we shall show later, it is the unimodal-distribution assumption, not the continuous nature of the memory trace that is contraindicated by the data. The a posteriori probability results for pairs are incompatible with the two-valued strength model, but this incompatibility is just the kind that could result from averaging individuals. Hence we must be cautious in deciding that the memory trace has more than two values in either single-digit or pair recognition, though the group data are certainly not well described by the two-valued strength model in either case.

Although we cannot conclusively establish whether the memory trace is two valued, multivalued, or continuous for either single digits or pairs, we can reach a definite conclusion about the process of incrementing the strength of digit pairs upon presentation of the pair: Some pairs are not incremented by presentation. With a two-valued or multivalued strength model, this amounts to assuming that presentation increments the strength of a pair with some probability, $\pi < 1$. This probability cannot be unity because in that case there would be no old pairs in the category with the lowest strength. The deterministic-increment model predicts that the straight-line portion of the MOC curve would be parallel to the horizontal axis, running through the point (1, 1). This model is clearly wrong. With a continuous-strength model, the MOC curve for pairs clearly requires that the strength distribution for old pairs be bimodal, with one portion of the distribution being directly beneath the distribution for new pairs and proportional to it. The most reasonable way for this to occur is for some pairs not to be incremented in strength by presentation, which is a formal way of stating one consequence of not "attending" to a pair. Thus, whether trace strengths are two valued, multivalued, or continuous, it is clearly necessary to assume that pair strengths are incremented with probability, $\pi < 1$. On the other hand it is perfectly reasonable to conclude from the present data that the trace strength for a single digit is incremented with probability, $\pi = 1$. This result is quite plausible on intuitive grounds if we assume that *S*s code and rehearse the digits in nonoverlapping pairs. This would result in attending to every digit, but only about half of the pairs (since there are five items in a list). This intuitive notion is confirmed by the fact that the best estimate of π for ordered pairs is .42.

Two formal models can describe

the paired-digits results: a multivalue strength model and a continuous-strength model. Both must assume a probabilistic, rather than a deterministic, incrementing process. The multivalue strength model is very similar to the multistate threshold model developed by Norman (1964) for psychophysical-detection experiments and will not be presented here. The probabilistic increment, continuous-strength model has not been stated previously and, for this reason, we now present a brief summary of the properties of such a model.

Assume some initial distribution of strengths (s) for all new items (or pairs of items), $f_n(s)$. We make no assumptions of f_n except that it be a continuous, unimodal probability-density function, where s has a lower bound of 0, so that $f_n(s) = 0$, $s < 0$, and

$$\int_0^\infty f_n(s)ds = 1.$$

When an item is presented, its strength in memory receives some increment with probability π (and no increment with probability $1 - \pi$). Let the stochastic operator T reflect the transformation of new strengths into incremented (old) strengths. The probability distribution of the strengths of old items is then simply

$$f_0(s) = \pi T[f_n(s)] + (1 - \pi) f_n(s).$$

The probability of a response of i or less to an old item is

$$p(i|oo) = \pi \int_{c_i}^\infty T[f_n(s)]ds + (1 - \pi) \int_{c_i}^\infty f_n(s)ds \quad [1]$$

and of a response of i or less to a new item is

$$p(i|nn) = \int_{c_i}^\infty f_n(s)ds, \quad [2]$$

where c_i is the decision cutoff for this criterion.

Solving Equations 1 and 2 for the equation of the MOC curve yields

$$p(i|oo) = (1 - \pi)p(i|nn) + \pi \int_{c_i}^\infty T[f_n(s)]ds. \quad [3]$$

If the increment in strengths governed by the operator T is sufficiently large to preclude much overlapping of the new distribution with the incremented part of the old distribution, then for strengths between zero and the region where the overlap begins, the value of the integral in Equation 3 is unity, so that

$$p(i|oo) = (1 - \pi)p(i|nn) + \pi. \quad [4]$$

Equation 4 describes an MOC curve which is a straight line of slope $1 - \pi$ that passes through point (1, 1). This line describes the linear part of our curves and allows us to estimate π as .42 for ordered pairs.

The model can be made more specific by assuming that, with probability π, a presentation of an item causes its strength

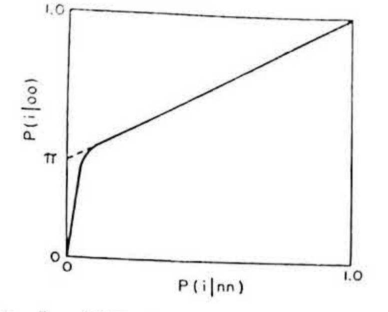


FIG. 5. MOC curve and theoretical distributions of trace strength for the probabilistic-increment, continuous-strength model. (The shaded area represents the old distribution, given by the sum of its two component distributions. An item, or pair, presentation is assumed to increment the strength of the trace in memory by a constant amount, Δ; the parameters are chosen to illustrate the ordered-pair data.)

in memory to be incremented by a constant amount of size Δ. In other words,

$$T[f_n(s)]=\begin{cases} f_n(s-\Delta), & 0\leqq(s-\Delta), \\ 0, & (s-\Delta)<0. \end{cases} \quad [5]$$

The model described by Equations 3, 4, and 5 is illustrated in Fig. 5A and 5B with the parameters chosen to illustrate the ordered-pair data.

This model has one important virtue over the multivalued strength model: It has the proper form for both the single- and paired-digit results. If we let π equal unity, the model contains two overlapping continuous distributions of the form required by the single-digit results. Further development and specification of the functions and operator cannot be performed with the limited data of this study.

REFERENCES

Bush, R. R., Galanter, E., & Luce, R. D. Characterization and classification of choice experiments. In R. D. Luce, R. R. Bush, and E. Galanter (Eds.), *Handbook of mathematical psychology*, Vol. 1, New York: Wiley, 1963. Pp. 77–102.

Egan, J. P., Schulman, A. I., & Greenberg, G. Z. Operating characteristics determined by binary decisions and by ratings. *J. Acoust. Soc. Amer.*, 1959, **31**, 768–773.

Luce, R. D. A threshold theory for simple detection experiments. *Psychol. Rev.*, 1963, **70**, 61–79.

Nachmias, J., & Steinman, R. M. Study of absolute visual detection by the rating-scale method. *J. Opt. Soc. Amer.*, 1963, **53**, 1206–1213.

Norman, D. A. Sensory thresholds, response biases, and the neural quantum theory. *J. math. Psychol.*, 1964, **1**, 88–120.

Pollack, I., & Decker, L. R. Confidence ratings, method reception, and the receiver operating characteristic. *J. Acoust. Soc. Amer.*, 1958, **30**, 286–292.

Swets, J. A., Tanner, W. P. Jr., & Birdsall, T. G. Decision processes in perception. *Psychol. Rev.*, 1961, **68**, 301–340.

Thurstone, L. L. A law of comparative judgment. *Psychol. Rev.*, 1927, **34**, 273–286.

(Received July 20, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 5, 490–495

ROLE OF REWARD MAGNITUDE AND INCOMPLETE REDUCTION OF REWARD MAGNITUDE IN THE FRUSTRATION EFFECT

JAMES H. McHOSE

Southern Illinois University

AND H. WAYNE LUDVIGSON

University of Texas Southwestern Medical School

30 naive food-deprived (22-hr.) male rats received 72 acquisition and 96 postshift trials in an L-shaped double-alley apparatus. In acquisition 16 Ss (Group 10) received 10 45-mg. pellets in the 1st goal box (G_1) on each trial, while 14 Ss (Group 2) received 2 pellets in G_1. In every block of 6 postshift trials, Group 2 received 2 pellets in G_1 on 3 trials and no reward on 3 trials, while Group 10 received 10 pellets on 3 trials, 2 pellets on 1 trial, and no reward on 2 trials. Both groups always received 2 pellets in the 2nd goal box. Final postshift speeds were inversely related to absolute (postshift) G_1 reward magnitude, but unrelated to expected (acquisition) reward magnitude.

A number of studies have shown that, after sufficient training in the double-alley apparatus, rats run faster in the second alley (A_2) on a certain trial if they have not been rewarded in the first goal box (G_1) than if they have been rewarded there. This difference in A_2 response vigor, termed the frustration effect (FE), is the basic datum underlying the hypothesis that, when an organism has acquired an "expectancy" of reward (r_g), nonreward is "frustrative" (Amsel, 1958, 1962). That is, there results a reaction of primary frustration (R_F) which contributes an increment to the drive level, and, in the double-alley apparatus, augments response vigor in A_2. The present study was, in part, concerned with the question of whether a reduction in reward to nonzero magnitude would, like a reduction to zero, also produce frustration, i.e., facilitate A_2 performance. In addition, this investigation was designed to determine whether the amount of response facilitation (frustration) produced by reward removal would depend upon the magnitude of G_1 reward obtained on reinforced trials.

In connection with the present study, it is important to note that, by itself, the FE is not conclusive evidence that reward removal per se is an additional source of drive. The FE could as well result from the removal, on nonrewarded trials, of a response-depression factor associated with reinforcement. The critical evidence for the frustration-drive hypothesis comes, then, from studies (McHose, 1963; Wagner, 1959) which demonstrate that nonreward produces a more vigorous response in A_2 when there has been a history of reinforcement in G_1 than when there has been no prior reward in G_1. These studies support the hypothesis by showing that the FE is, at least in part, a relative effect, resulting from *removal* of reward. And they rule out the simpler alternative that the FE results from a response depression which is an increasing function of the absolute amount of reward received immediately prior to A_2 performance. Any study which purports to demonstrate a frustration-drive phenomenon resulting from a change in reward magnitude must, on a given trial, show the effect of the change over and above any effects of the absolute quantity of reward received on that trial. In order to demonstrate, for example, that a reduction from 10 to 2 pellets results in a frustration-drive increment, it must be shown that rats trained on 10 pellets of reward run faster following a reduction to 2 pellets than rats which have received only 2 pellets from the outset of training.

A recent study which neglected to separate the effects of absolute magnitude from degree of reward reduction was, like the present study, concerned with the question of whether incomplete reduction of reward, i.e., reduction to nonzero magnitude, would also augment A_2 response vigor. Bower (1962) reported that *S*s trained on a relatively large reward in G_1, and subsequently given test trials with reward reduced by varying amounts, show faster A_2 test-trial speeds the smaller the reward in G_1. Although this finding could reasonably be attributed to a graded frustration effect resulting from increased reduction of reward, the within-*S* design of the experiment does not preclude the possibility that some or all of the effect results from a graded depression of response speed produced by absolute reward magnitude. Thus similar test-trial speeds might have been obtained regardless of the magnitude of the training reward and the amount of reward reduction. The present study explored this possibility.

The present study also tested the implication of frustration theory that the magnitude of the frustration-drive increment increases with the magnitude of the expected or training reward. Such an effect would be expected since the intensity of R_F is assumed to increase with reward expectancy, r_g, and r_g, in turn, is assumed to increase with reward magnitude. For two groups trained on different reward magnitude in G_1 and then given test trials with the training magnitude and with no reward, a differential drive increment would be indicated by a difference in A_2 speeds on nonrewarded trials. It should be emphasized that the difference must appear on nonrewarded trials, since it is only on these trials that speeds are assumed to be elevated by frustration drive.

METHOD

Subjects.—The *S*s were 30 experimentally naive, male, hooded and albino rats ranging in age from 90 to 100 days at the beginning of the experiment.

Apparatus.—The apparatus was an L-shaped double runway comprised of a start box (S_1), alley (A_1), goal box (G_1 or S_2), and a second alley (A_2) and goal box (G_2). The first goal box served as a start box for the second alley. The start and goal boxes were 12 in. long and the alleys 24 in. long, with a 4-in. height and 3-in. width throughout. Sections S_1, A_1, and G_1 were painted flat black, while Sections A_2 and G_2 were painted white. All segments of the apparatus were covered with clear glass. Wooden blocks with indentations for food pellets served as goal cups in G_1 and G_2.

Sliding start doors separated S_1 from A_1 and S_2 from A_2, and guillotine doors which prevented retracing were located between A_1 and G_1, and between A_2 and G_2. Opening the second start door activated a 1/100-sec. clock which stopped when *S* interrupted a photocell beam located 6 in. from S_2. For half of the *S*s, interruption of this photocell beam started a second 1/100-sec. clock which stopped when *S* crossed a second photocell beam located 18 in. from S_2.

Procedure.—Twenty days prior to the first training day, *S*s were placed on a 23-hr. food-deprivation cycle which was maintained throughout the experiment. Purina lab chow was available in the home cages for a 75-min. period each day. On each of the 4 days prior to the first training day, *S*s were fed approximately 1 gm. of 45-mg. Noyes pellets, identical with the pellet used throughout training, before receiving their daily ration. Water was always available in both home and carrying cages.

The experiment consisted of a 72-trial acquisition period followed by a 96-trial postshift phase. During acquisition each of 16 *Ss* (Group 10) received 10 pellets in G_1 and 2 pellets in G_2 on each trial, while 14 *Ss* (Group 2) received 2 pellets in G_1 and 2 pellets in G_2. In every block of six postshift trials, Group 2 received 2 pellets in G_1 on three trials (R trials) and no reward on three trials (0 trials), according to the following repeating cycle: ROROROORORO-RROROORORORRO. In each postshift block of six trials, Group 10 received 10 pellets in G_1 on three trials (R trials), 2 pellets on one trial, and no reward on two trials, according to the following repeating cycle: ROR2ROOROR2RR2ROOROR2RROROR-OR22ROROR. Both groups continued to receive 2 pellets in G_2 on each trial. The *Ss* received six trials per day, with the exception of the first 2 days of training, on which three trials were given. Within any day the intertrial interval was approximately 10 min. The experiment was conducted in two replications with 8 and 7 *Ss* randomly assigned to Groups 10 and 2, respectively, in each replication.

A trial began with the introduction of *S* to S_1. The first start door was opened after *S* had maintained an orientation toward it for 3 sec. When *S* entered G_1, the retrace door separating A_1 and G_1 was closed. The second start door was opened after *S* had eaten and maintained an orientation toward it for 3 sec. On nonrewarded trials, the second door was opened after a 3-sec. orientation period. When *S* entered G_2, the retrace door separating A_2 and G_2 was closed and *S* was removed immediately after eating. All *Ss* were fed after the last experimental trial for the last *S* of the day.

Second-alley performance was recorded in terms of reciprocal traversal times over the first 6 in. of A_2 (starting speeds) in both replications. In the second replication, reciprocal traversal times over the next 12 in. of A_2 (running speeds) were also recorded.

In analyzing the starting-speed data, replications and all interactions with replications were evaluated as separate sources of variance. However, these effects proved generally negligible and they were pooled into the appropriate error terms. Since none of the conclusions from statistical tests was thereby altered, only the latter analyses will be discussed.

RESULTS

Figure 1 presents group mean Alley 2 starting speeds as a function of trial blocks. The vertical line at Trial Block 12 indicates the point at which the reinforcement schedule in G_1 was changed. Each preshift block is composed of 6 trials. The postshift data were separated into blocks of 12 trials. Within each postshift block for Group 10, mean speeds for the 6 trials on which *Ss* received 10 pellets (10-10), the 2 trials on which they received 2 pellets (10-2), and the 4 trials on which they received no reward (10-0) in G_1, are plotted separately. Similarly, the mean speeds for the 6 trials on which Group 2 received 2 pellets in G_1 (2-2) are plotted separately from those for the 6 trials on which they received no reward in G_1 (2-0).

Simple analysis of variance performed over Trial Blocks 10–12 showed the apparent superiority of Group 2 in acquisition to be statistically reliable, F (1, 28) = 7.99, $p < .01$.

Inspection of Fig. 1 indicates that, for Group 10, postshift speeds increased as the amount of reward obtained in G_1 decreased. Analysis of variance on the Group 10 data at Blocks 13 and 20 indicated statistical significance for the two main effects, amount of reward, F (2, 30) = 65.39, $p < .001$, and trial blocks, F (1, 15) = 23.55, $p < .001$, as well as their interaction, F (2, 30) = 4.77, $p < .025$.

Comparison of the 10-10 and 10-0, and the 2-2 and 2-0 trend lines, suggests that the A_2 speeds of both groups were faster on nonrewarded trials than on rewarded trials, with this effect being more pronounced

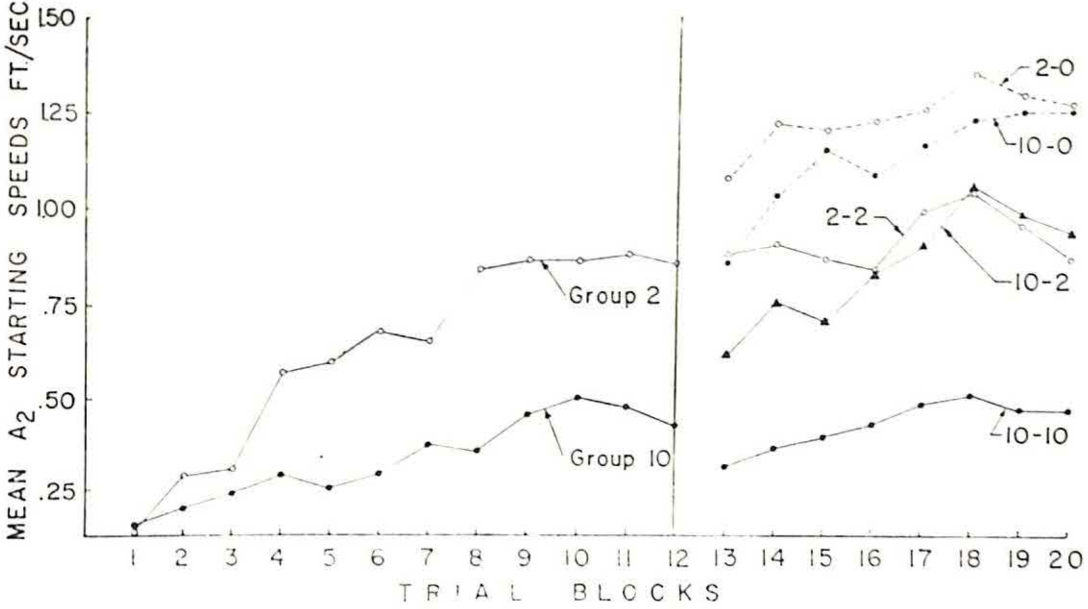


FIG. 1. Alley 2 starting speeds as a function of G_1 reinforcement conditions.

in the Group 10 data. A three-dimensional analysis of variance appropriate for a design with one between-group and two within-group variables (reward or nonreward and trial blocks) was performed over Blocks 13 and 20. The analysis yielded a significant main effect of the reward-nonreward variable, F (1, 28) = 134.63, $p < .001$. The Reward-Nonreward × Groups interaction was also significant, F (1, 28) = 32.65, $p < .001$, indicating a greater reward-nonreward speed differential for Group 10. Subsequent analysis within this design of the data for both groups on nonrewarded trials, i.e., 10-0 and 2-0 data, yielded a significant overall trials effect, F (1, 28) = 22.95, $p < .001$. The Trials × Groups interaction was not significant, failing to support the graphical impression of convergence of the two trend lines.

A comparison of the 10-2 and 2-2 trends indicates an apparent convergence in the postshift period. However, the Groups × Trials interaction over Blocks 13 and 20 was not significant. The overall trials effect was significant, F (1, 28) = 30.54, $p < .001$.

Finally, comparisons of Group 10-0 with 2-0, and 10-2 with 2-2, at Block 20 yielded F ratios less than unity, supporting the graphical impression that the asymptotic speed resulting from a given G_1 reward was independent of the relationship of that reward to the expected reward.

Although the running-speed data, obtained in one replication only, proved less statistically reliable, they were in close agreement with the starting-speed data in both the acquisition and postshift periods, with one exception. In the running-speed data, the final performance level of Group 10 on nonrewarded trials approximated that of the 10-2 and 2-2 conditions.

DISCUSSION

The postshift data for Group 10, consonant with those of Bower (1962), clearly show that A_2 test-trial speed increases as G_1 reward decreases. However, the present data do not support the view that this effect is a graded frustration-drive phenomenon attributable to increased reduction in reward from an expected (preshift) level to obtained postshift levels. This is apparent from inspection of the 10-2 and 2-2 data over the latter trials of the postshift phase. Here *S*s trained in the preshift period to expect a large reward perform no differently from *S*s trained to expect a small reward in spite of the substantial reduction experienced by the former. This indicates that these A_2 speeds are simply a function of absolute reward magnitude in G_1, and that no frustration-drive increment from incomplete reward reduction need be invoked.

Looking next at the influence of expected reward magnitude on the standard FE measure, it may be seen that the difference between speeds on rewarded and nonrewarded trials is clearly greater for Group 10, i.e., the difference between the 10-0 and 10-10 curves is greater than that between the 2-0 and 2-2 curves. Of primary interest, however, is the fact that in the latter stages of the postshift period the difference in FE magnitude results from differential speed on rewarded trials, rather than nonrewarded trials. This indicates that the critical variable determining the size of the FE is, again, absolute reward magnitude and not amount of reward reduction. Therefore, frustration drive is irrelevant to the explanation of these data also.

It is important to note that these findings, though not explained by Amsel's (1962) formulation, do not vitiate the basic assumption of frustration theory, viz., that complete removal of an expected reward produces an increment to the drive level. The 10-0 and 2-0 speeds of the present study may be elevated by this increment, and, in fact, previous studies (McHose, 1963; Wagner, 1959) have shown the necessity of assuming such an elevation. What the present findings *do* indicate is that this drive increment does not occur with a substantial, but incomplete, reduction in reward, nor does it vary with reward magnitude as defined by number of pellets.

Since magnitude of reward is postulated (Amsel, 1962; Spence, 1956) to be a principal determinant of the strength of reward expectancy, r_g, these data imply that R_F and r_g do not covary in the manner heretofore believed. A theoretical difficulty posed by this conclusion stems from studies suggesting that other variables also assumed to determine r_g do influence the magnitude of R_F (Amsel, Ernhart, & Galbrecht, 1961; Amsel & Hancock, 1957). However, before it is concluded that these several variables do not all determine the same theoretical construct, r_g, the precise nature of the findings of these previous studies might be questioned. For the most part, these studies employed the standard FE as their measure of response elevation, and the difference between FEs as indicative of the effect of an expectancy-influencing variable. But, as in the present study, the difference between groups on nonrewarded trials is the critical comparison for a frustration-drive interpretation since speeds on rewarded trials may not be independent of the experimental manipulations. More importantly, if the expectancy-influencing variable is suspected of having an additional effect which would make speeds on nonrewarded trials differ because of some factor other than frustration, then the magnitude of the frustration-drive increment must be assessed by comparing speeds on nonrewarded trials against speeds for *S*s under identical conditions except for having had no prior reward in G_1. It should be noted that other studies have also failed to provide these critical comparisons (e.g., MacKinnon & Amsel, 1964; McHose & Ludvigson, 1964) and may therefore suffer an ambiguity in interpretation. In this connection, the present study points up the danger in applying the label "frustration effect" to the difference in A_2 speed on rewarded and nonrewarded trials for a single group of *S*s. Since it must be empirically demonstrated through the use of appropriate control groups that the FE does indeed index a frustration-drive increment, it would seem wise to reserve the term FE for the speed differential obtaining between *S*s experiencing reward reduction and their control *S*s.

While portions of the present data are either inconsistent with or unexplained by Amsel's drive-increment hypothesis, alternative formulations are also less than satisfactory. Bower's (1962) r_g-perseveration hypothesis, which proposes to account for the FE by assuming that the rate of decay of conditioned excitement from A_1 is directly related to the magnitude of the G_1 reward, fails because it also predicts that A_2 speeds are a function of expected as well as obtained reward. Since Bower's hypothesis is an extension of Spence's (1956) theory of incentive motivation, it implies that Group 10 should have a larger r_g amplitude conditioned to A_1 cues than should Group 2 because a larger G_1 reward was employed in acquisition training. Therefore, on test trials with zero G_1 reward, the strength of the perseverating r_g, and consequently A_2 speed, should be greater for Group 10. Since a similar argument holds for the two-pellet test trials, this hypothesis is directly contradicted by the present data.

The drive-reduction hypothesis (Seward, Pereboom, Butler, & Jones, 1957), that larger G_1 reward magnitudes produce greater hunger reduction and consequently slower speeds in A_2, most readily fits the present incomplete reward-reduction data. However, the scope of the drive-reduction hypothesis would be quite limited since it cannot account for the previously demonstrated effects of *complete* reward reduction (McHose, 1963; Wagner, 1959), effects which presumably elevate the 10-0 and 2-0 speeds of the present study.

REFERENCES

AMSEL, A. The role of frustrative nonreward in noncontinuous reward situations. *Psychol. Bull.*, 1958, **55**, 102–119.

AMSEL, A. Frustrative nonreward in partial reinforcement and discrimination learning. Some recent history and a theoretical extension. *Psychol. Rev.*, 1962, **69**, 306–328.

AMSEL, A., ERNHART, C. B., & GALBRECHT, C. R. Magnitude of frustration effect and strength of antedating goal factors. *Psychol. Rep.*, 1961, **8**, 183–186.

AMSEL, A., & HANCOCK, W. Motivational properties of frustration: III. Relation of frustration to antedating goal factors. *J. exp. Psychol.*, 1957, **53**, 126–131.

BOWER, G. H. The influence of graded reductions in reward and prior frustration events upon the magnitude of the frustration effect. *J. comp. physiol. Psychol.*, 1962, **55**, 582–587.

MACKINNON, J. R., & AMSEL, A. Magnitude of the frustration effect as a function of confinement and detention in the frustrating situation. *J. exp. Psychol.*, 1964, **67**, 468–474.

MCHOSE, J. H. The effect of continued nonreinforcement on the frustration effect. *J. exp. Psychol.*, 1963, **65**, 444–451.

MCHOSE, J. H., & LUDVIGSON, H. W. Frustration effect as a function of drive. *Psychol. Rep.*, 1964, **14**, 371–374.

SEWARD, J. P., PEREBOOM, A. C., BUTLER, B., & JONES, R. B. The role of prefeeding in an apparent frustration effect. *J. exp. Psychol.*, 1957, **54**, 445–450.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

WAGNER, A. R. The role of reinforcement and nonreinforcement in an "apparent frustration effect." *J. exp. Psychol.*, 1959, **57**, 130–136.

(Received July 13, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 5, 496–505

EFFECT OF SIZE AND LOCATION OF INFORMATIONAL TRANSFORMS UPON SHORT-TERM RETENTION [1]

MICHAEL I. POSNER [2] AND ELLEN ROSSMAN [3]

University of Wisconsin

This series of studies investigates the effect of informational transformations of various levels of difficulty, as indicated by the amount of information reduction required, upon material in short-term storage at the time of the transform. Experiment I shows that with number and similarity of interpolated items held constant, the greater the difficulty of a transform the more forgetting will result from it. Experiments II and III show that these effects cannot be attributed entirely to increases in the time an item remains in store. Rather, time in store and difficulty of transform both contribute to determining the amount of forgetting. Experiment IV shows that the loss of material in store is a decreasing function of its distance prior to the transform, but that the transformed material itself shows no decline in retention. These results point to an operational definition of rehearsal as a process requiring part of the limited central capacity of S.

Many sequential tasks involve both the storage of information and continuous transformations of it through classifications or combinations. Thus, reading a book or listening to a lecture combines a storage function with operations upon the input information. In the analysis of such tasks two lines of psychological research converge. The first involves the role of immediate memory in sequential tasks (Posner, 1963) and the second involves the measurement of difficulty of operations upon input information (Posner, 1964b).

Human performance in a sequential task resembles a standard computing system in that information is temporarily stored during the performance of transforming operations (Simon & Kotovsky, 1963). However, in comparison of human information processing with that of computers a key difference is in the effect of computing operations upon material in store at the time of the operation. A standard computing system is designed to preserve information which may be needed in future operations so as to avoid the arbitrary loss of material due to the transformations occurring in the units which operate upon information. The human being is surely more susceptible to loss of information, but the laws governing such loss are not fully understood. The study of the interaction of immediate memory and transformations is, therefore, important in any attempt to specify precisely the information-processing characteristics of the human.

From the point of view of memory two factors about any event which intervenes between a stimulus and its recall are known to be crucial. First, is the similarity of the interpolated event to material in store (McGeoch & Irion, 1951). Second, is the length of time which the interpolated event interrupts rehearsal of stored information (Brown, 1958). In addition some investigators (Pillsbury & Sylvester, 1940) have held that the difficulty of the interpolated activity effects retention independently of the other factors. Recently Melton (1963) has suggested that a more extensive analysis of tasks intervening between presentation and recall needs to be made in order to determine the relative importance of these various factors.

---

[1] The research reported in this study was supported by a grant from the Research Committee of the University of Wisconsin with funds provided by the Wisconsin Alumni Foundation. The paper was prepared while the senior author was Visiting Research Psychologist at the Human Performance Center, Department of Psychology, University of Michigan, supported by the Advanced Research Projects Agency of the Department of Defense and monitored by the Air Force Office of Scientific Research, under Contract No. AF 49(638)-1235.

[2] Now at the University of Oregon.

[3] Now at the University of California, Berkeley.

Posner (1962, 1964b) has presented a method for the a priori prediction of the difficulty of informational transformations in terms of the difference between stimulus and response information (information reduced). Many tasks which require transformation of input involve information reduction. The tasks falling within this general category may differ rather widely. However, within relatively restricted task configurations the amount of information reduced appears to be a quantitative predictor of task difficulty. Specifically for a set of simple numerical tasks with the same stimulus input it was found that errors increased and rate of processing decreased as the amount of information reduction required by the task increased. Six different tasks were used including recording, addition, and classification. For this set of tasks the amount of information reduction is directly related to task difficulty. The question of the generality of this relation between information reduction and difficulty for tasks other than those which have been studied specifically will be discussed in the conclusion section of this paper.

This series of studies is designed to investigate the interrelation of short-term memory and information transformations. The goal is threefold. First, to determine the effect of increasing the difficulty of a transform upon material in store at the time of the transform. Second, to determine if such effects can be completely attributed to similarity of interpolated material and/or length of time between presentation and recall. Finally, to determine the location of any increase in forgetting resulting from the transform.

EXPERIMENT I

The purpose of this study was to investigate the effect of number and size of informational transforms upon the amount recalled. In addition, it provided the opportunity to confirm the previous result that the difficulty of a transform can be predicted from a knowledge of its informational properties.

Method

Subjects.—The *S*s were 53 students in introductory psychology who received course points for their work. The data of 5 *S*s were not considered; 3 failed to follow the required instructions as to the digits to transform, 1 was given an improper sequence, and 1 was not completed due to equipment difficulties.

Material.—Five practice and 40 regular series of 8 digits were recorded at a rate of 30 digits/min. Each digit was selected at random with the restriction that no two successive digits be identical. There was a 20-sec. pause between series. The digits were recorded over white noise which served to mask any disturbance from outside the experiment.

Tasks.—The four types of informational transforms interpolated were: (*a*) Reversal: In this task *S* was required to write down a pair of digits in the opposite order from their presentation. This transformation had the same input and output information and was thus 0 bits reducing. (*b*) Addition: In this task *S* was required to add two adjacent digits together and record the sum. Input information for the pair was about 6.6 bits, while the sum contains only 3.8 bits. The task therefore requires 2.8 bits reduction. (*c*) 2-Bit classification: In this task *S* was required to classify each pair of numbers into high (50 and above) and low (below 50) and into odd or even. The input information is the same as in previous tasks while the output is 2 bits, yielding a 4.6-bit reduction. (*d*) 1-Bit classification: In this task *S* records "A" if the pair of numbers is high and odd or low and even, and "B" for the reverse. Input is again 6.6 bits and output 1 bit, yielding a 5.6-bit reduction.

Procedure.—The *S*s were randomly assigned to one of four groups, with 12 *S*s in a group. Each group performed only a single task. The *S*s received instructions on how to handle the pairs which they were asked to transform. They were provided with a score sheet which was folded in two. On the left side were eight boxes, each of which contained the letters H, L, O, E in the upper portion and the letters A and B in the lower portion. The *S*s in the reversal group and the addition group wrote their transformation in the square corresponding to the second digit of the pair(s), which they were instructed to transform, *S*s in the 2- and 1-bit classification groups used the same square but drew a line through the letters representing the appropriate category. After each series of eight numbers was over *S*s turned over the sheet immediately and began to recall the numbers they had received, writing them in the eight boxes provided. The *S*s were instructed to record only in the order of presentation, always starting with the earliest digit they could recall.

All *S*s began with five practice problems. The first two practice series required no transform, then transformation of the last pair, the last two pairs, and the final series required transforms of the last three pairs. Following this *S*s received the 40 regular series. Before each set of 10, *S*s were told whether to classify no pair, last pair, last two pairs, or last three pairs. The order of these tasks was balanced by three random Latin squares within each group.

Results

Difficulty of the transforms.—Since the rate of presentation of the materials in this situation was slow there were almost no omitted transforms. Each task involves a different chance probability so that the percentage of correct transformations above the chance level was calculated for each task. The percentages were 99, 94.8, 80.6, and 70.3 for the reversal, addition, 2-bit, and 1-bit tasks. If the percentage correct is plotted against the information reduction the result is a strong linear trend. While direct comparison of tasks with very different chance levels is risky, this result corresponds quite well to the relation previously found between performance and information reduction when similar tasks were speeded and the performance decrement for each task measured as rate increased (Posner, 1962). Thus it appears that the a priori information-reduction measure is related to task difficulty.

Amount retained.—To determine the effect of the transforms upon recall, the number of errors in the first two digits of each series was determined for each *S*. This measure was selected since the first pair is never itself transformed or recorded and is always recalled first by *S*. Thus the only difference between treatments with respect to the first pair of digits is the nature of the intervening transformations. Table 1 shows the total errors on the first pair of digits for all conditions and all numbers of transforms.

A trend analysis was performed on these data to determine the form of the function which would best describe the relationship between errors and number of transforms for each task. The analysis indicates that errors vary linearly with number of transforms for each task ($p < .01$). In addition a significant Linear Task × Transform interaction ($p < .01$)

TABLE 1

Total Errors in the First Pair of Digits

Task	No. of Pairs Transformed				
	0	1	2	3	Total
Reversal	22	26	31	50	129
Addition	44	62	56	104	266
2/Bit	29	50	71	114	264
1/Bit	28	62	94	108	292
Total	123	200	252	376	

Note.—Each cell represents 120 possible errors.

indicates that the slopes of these functions are different for the various tasks.

The significant Task × Transform interaction means that the effect on retention of increasing the number of transforms is different depending upon the task. It is possible to obtain an estimate of the amount of forgetting caused by a transform of various sizes. To do this the number of errors per transform is averaged for the 1, 2, and 3 transform conditions after subtracting out the number of errors made by that group with zero transforms. This is done for each of the tasks. The average increases in errors per transform for the 0-, 2.8-, 4.6-, and 5.6-bit reduction tasks are 5.9, 14.7, 23.4, and 31.2, respectively. Comparing the slopes of best-fitting lines relating number of transforms and errors gives substantially the same results except that the 4.6- and 5.6-bit reduction tasks are about equal. This is because there is a relatively small difference between two and three transforms in the 5.6-bit task and this has a large effect on the overall slope.

The general result is clear. Both the number of transforms and their size have significant effects upon the amount of material lost from store. The amount of forgetting increases linearly with the number of transforms, the average increase being 15.7 errors per transform. The amount of forgetting also increases roughly linearly with the size of the transform (at least up to 4.6 bits reduction) so that on the average the additional errors caused by adding a single transform run from 4.9 with the 0-bit reduction task to 25.9 with the 5.6-bit reduction task.

Effect of similarity.—In each of these tasks the interpolated digits are identical so that there are no differences in the interpolated stimuli. The products of the transform are, however, different. In two cases they are numbers and in two cases letters representing categories. The reversal and addition tasks have transformed products of greater over-all similarity to the stored material than the classification tasks. It is clear that the greater forgetting found in the classification tasks cannot, therefore, be due to increased similarity. In fact these results are in accord with many findings in the area of immediate memory that similarity of interpolated material is not an important factor in the amount of forgetting (Posner, 1963).

Errors in all pairs.—The total errors in recall of the series is a somewhat confounded measure. If, for example, *S* remembers the product of the reversal transform he can, by making another reversal, produce the correct recall. The size of the transform is inversely related to the utility of its product in recalling untransformed information. Despite these difficulties the results of an analysis of errors in all positions are not unlike that presented for the first position. A linear relation between errors and number of transforms is found for each task. The increase in Errors/Transform is 20.1, 49.7, 76.2, and 90.7

for the 0-, 2.8-, 4.6-, and 5.6-bit tasks. The relation is linear, much like that found for the first pair.

EXPERIMENT II

The results of Exp. I indicate that transforms do interfere with retention and that such interference is more closely related to the size and number of transforms than to the similarity of the transformed product. However, Posner (1962) has shown that the size of a transform is related to the length of time required to perform it, so that it is possible that these effects are due to the length of time required for the transform. It is also possible that the effects of interpolated tasks would be unequal even for a fixed period of time. That is, whatever process sustains the memory may be able to coexist to different degrees with differing tasks. This study is designed to investigate this possibility by keeping the total interpolated time constant but varying the size of the transforms performed in that time.

Method

Subjects.—The 28 *S*s were obtained from the same source as in the previous study. Three *S*s were eliminated for failure to follow instructions in the pairs transformed.

Material.—The material consisted of 50 series of 5 digits recorded over white noise and presented at a rate of about 2/sec. After the burst of 5 digits there was 10 sec. of white noise alone and then the word "recall," followed by 20 sec. of white noise alone. Each *S* was provided with a set of score sheets which contained digits to be used in the interpolated task.

Tasks.—Two of the interpolated tasks were exactly the same as in Exp. I. These were addition and 2-bit classification. Three new tasks were added: (*a*) Control: No interpolated task. (*b*) Recording: The *S* copied the number without change. This task is 0 bits reducing. (*c*) Counting backward: The *S* counted backwards by threes from the last two numbers of the tape, recording each successive result. The difficulty of this task had not been studied previously but it has been used frequently in memory studies (Peterson & Peterson, 1959). It seemed useful, therefore, to consider it in relation to the other tasks. It is possible to infer the relative difficulty of this task from the rate and accuracy of the transforms actually made by *S*s; however, the information reduction required by this task was first estimated in the following way. Counting by one may be considered to present all the information in a sequence of numbers. If numbers are skipped then this requires selection of a part of the information in the series. Since *S* is dealing with a rapidly decreasing population of numbers from 0 to 99 we consider that the average population of numbers dealt with over the task is 50. Thus counting by one generates $\log_2 50 = 5.64$ bits and selection of one third of this yields 1.88 bits, a reduction of 3.8 bits.

Procedure.—Each *S* performed five practice series, one using each of the tasks. He then performed 50 regular series, 10 of each task, determined by a random Latin square which was replicated five times. For each task *S* listened to the five numbers, using the last two to begin his interpolated task. After processing the pair on the tape, *S* was instructed to shift to the score sheet and process as many pairs as he could until the word "recall" sounded. Subjects then recalled the *first three digits* in the order presented and prepared for the next series.

Results

The percentage of series in which there was an error was computed for each interpolated task and averaged over the 25 *S*s. These were .4, 12.4, 21.6, 34, and 44.8 for the control, record, addition, backward, and classification tasks, respectively. Figure 1 shows the percentage of series in error (solid line) and the percentage of individual items in error (dotted line) plotted against the amount of information reduction required by each of the tasks except the control. Both scores show a regular increase with amount of information reduced. Analysis of the individual and combined Latin squares showed highly significant ($p < .01$) effects of treat-

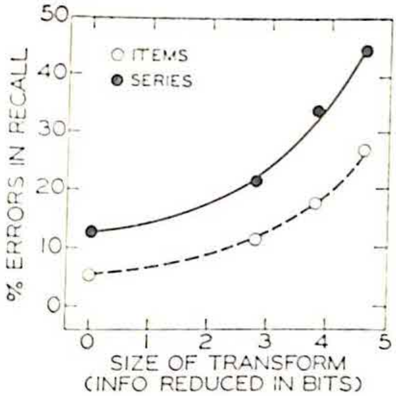


FIG. 1. Percentage of recall as a function of size of interpolated transforms.

ments and *S*s. The effect of order was significant ($p < .05$) due to a small drop in errors over successive tasks which was most apparent in the final 10 series. Since *S*s handled only 10 series of each type, learning effects were, as expected, rather modest. There were differences between Latin squares ($p < .05$) but these did not interact with conditions so that it seems reasonable to consider the grouped data. In fact each individual Latin square revealed the identical significant effect of treatments.

Two of the tasks involved no information reduction. In the record task, however, *S* read and wrote an average of 11 digit pairs during the 10-sec. pause, while in the control condition only the two digits on the tape intervened. The effect of having such a large number of interpolated items is to cause a relatively small increase in forgetting. The differences here are nowhere near as great as between the record and classification conditions although in classification only about half as many digits are interpolated as in record. These findings demonstrate that the effects of task difficulty as measured by amount of information reduction are at least as potent as those of number of interpolated items and length of time, the traditional variables considered in memory work.

Transform difficulty.—That *S*s actually found the interpolated tasks of varying degrees of difficulty is shown by the number of pairs of digits actually handled during the interpolated task. This value is 5.2, 5.5, 11.8, and 11.3 for the classification, backward, addition, and record tasks. While there was very little difference between the classification and backward tasks and between the add and record tasks in number of transformations performed, the percentages of errors were 1.8, 1.3, .61, and .18 for the four tasks. Thus regardless of whether the task is self-paced, as in this study, or *E*-paced as in the previous study, the number of errors per operation increased uniformly with the amount of information reduced. It is particularly interesting that the backward task lies about halfway between classification and addition just as the a priori informational analysis and the analysis of recall errors would suggest.

EXPERIMENT III

In this experiment the effect of varying time in store for four levels of the interpolated task difficulty is examined.

Method

Subjects.—The 23 *S*s were obtained from the same source as in Exp. I. Three *S*s failed to follow instructions and are not included.

Material.—The materials were identical with those used in Exp. II except that there were now 80 series divided into four sets of 20 each. Within each set there were supposed to be 5 series each, with 5-, 10-, 20-, and 30-sec. delay between presentation and recall. Due to an error, the third set contained one extra series at 30 sec. and one less at 20 and the fourth set contained one extra at 10 and one less at 20. The order of the various delays was randomized within each set. In addition there were four practice series, all with 10-sec. delay.

Procedure.—Each *S* performed the four practice series and each of the 80 regular series. Each *S* used four interpolated tasks, record, addition, backward counting, and classification (see Exp. II), one for each set of 20 series. The order of performing the tasks was controlled by a 4 × 4 Latin square. Each four *S*s comprised a replication of the study so that in all five random Latin squares were used. The design insures that each task was performed an equal number of times with each of the number sets; however, a particular series was associated with only one delay interval.

Results

The number of correct items recalled in each three-item series was scored. There were 16 experimental conditions in all (Four Tasks × Four Delays). Because of the error in the number of series at each delay interval the total items over all 20 *S*s which were presented in the four delay conditions were not identical. For each interpolated task at 10 and 30 sec. there were 315 possible items, 300 at 5 sec., and only 270 at 20 sec. These values serve as the denominators for calculating the percentage of correct items in each condition. Figure 2 shows the percentage of correct items as a function of time in store for each of the interpolated tasks. Reading the four curves from bottom to top they represent the record, add, backward counting, and classification tasks, respectively. Looking first at the 10-sec. delay condition the points represent a replication of Exp. II. As can be seen the essential features of the curve for items shown in Exp. II are duplicated in this study.

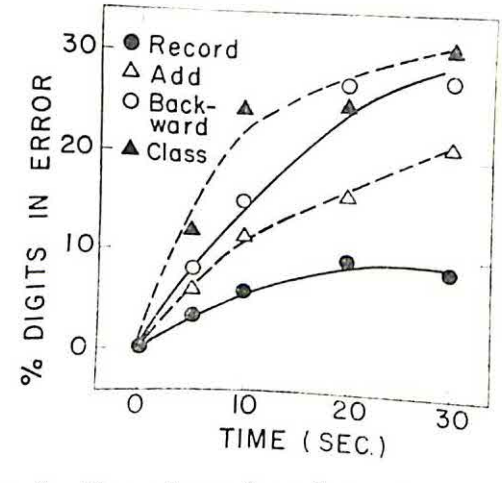


FIG. 2. Percentage of recall as a function of time for various size transforms.

The average percentages of errors at the four delays were: 7.7, 13.3, 19.4, and 22 for the 5-, 10-, 20-, and 30-sec. conditions. The control condition for Exp. II showed that no errors were to be expected with these materials at 0 delay. The magnitude of the effects due to task difficulty is about the same as for the particular intervals of delay used since the average percentages of errors are 5.92, 13.8, 19.2, and 22.5 for the record, add, backward counting, and classification tasks, respectively.

EXPERIMENT IV

In the previous experiments attention has been focused on the effect of the size of a transform upon forgetting. In considering sequential tasks another important aspect of a transform is its location within the series. In this experiment the essential technique of Exp. I is used but with the size of the transform held constant and its location within the series varied.

Method

Subjects.—The 13 new *S*s were obtained from the same source as in Exp. I. One *S* failed to follow the instructions and his data were not considered.

Procedure.—All *S*s performed the 2-bit classification task. The materials and apparatus were the same as in Exp. I. Each *S* was given five practice series, one each with no transforms, last pair, third pair, second pair, and first pair. He then received 40 regular series, 10 classifying each of the four pairs of the series. The order of presentation was controlled by three random Latin squares. Otherwise the procedure was exactly as in Exp. I.

Results

The results of this experiment appear in Table 2. The total number of errors which occurred with classification of the first through fourth pair were 155, 250, 224, and 253. Table 2 shows the percentage of errors for each pair of the series. To see the effect of transforms on material in store a standard serial-position curve obtained from 10 different *S*s over the same material is included (Posner, 1964a). When the first pair is classified, the serial-position curve appears very similar to the no-classification condition. When, however, the second pair is classified the percentage of errors in the first pair is more than doubled. This effect is displayed graphically in Fig. 3. Over all conditions the percentage of errors which occurred at each pair is averaged according to the distance from the transform. The dotted line represents the normal serial-position curve for no transforms. The line marked −1, e.g., includes the percentage of errors in the first pair when the second pair was transformed and the percentage of errors in the second pair when the third pair was transformed, etc. The line marked T includes the errors at the transformed position itself. Points at +1, −2, and −3 represent one position after the transform, two before, and three before the transform.

TABLE 2

PERCENTAGE OF TOTAL ERRORS BY SERIAL POSITION

Transform Position	Serial Position			
	1	2	3	4
None	16.5	28.2	35.7	19.6
1	11.0	25.8	41.3	21.9
2	26.4	24.0	32.0	17.6
3	19.2	34.4	25.4	21.0
4	17.4	30.8	43.9	7.9

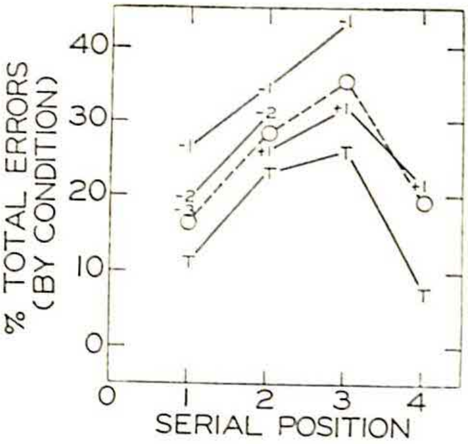


FIG. 3. Effect of a transform upon retention of material at various positions in the series.

The location of the transform within the series affects what is forgotten. In general, the material immediately prior to the transform is most affected and this decreases with distance. Forgetting of the transformed material itself is not increased nor does the material after the transform show a decrease in recall. If the transform starts the series much less effect is found than if it occurs at any of the other positions.

DISCUSSION AND CONCLUSIONS

The results of this study are consistent with a functional model of human information characteristics which has been evolved over the last several years (Broadbent, 1958). The results confirm many studies which show that retention in short-term memory is an active process which is extremely liable to disruption (Posner, 1963). If the human is considered as an information-processing system with limited capacity (Broadbent, 1958) these studies indicate that whatever process sustains the memory trace during the first few seconds after the stimulus requires a portion of that capacity. This process has often been called rehearsal. However, the tendency to use the term rehearsal as a covert analogue of speech has led to the view that it is either entirely present or absent. Thus an interpolated task is said to "prevent rehearsal" if it demands the attention of *S*.

There are a number of advantages replacing the notion of rehearsal as an all-or-none process with a view of *S* as having a limited capacity for information processing. The rehearsal process requires a part of this capacity and can coexist to a greater degree with tasks which require a small amount of this capacity than with those requiring a larger amount. Thus, Exp. I, II, and III show a close quantitative relation between the size of the transform and the amount of forgetting caused by it. They demonstrate that the amount of retention for a given period in store can vary greatly depending upon the processing capacity available to the rehearsal process. This view is closely related to the older idea of perseverative consolidation and to other views of autonomous enhancement (Melton, 1963), but is somewhat more limited since only processes which can be shown to require part of *S*'s central capacity can be said to involve rehearsal. This definition allows operational control of rehearsal by manipulation of the available processing capacity. Moreover this view makes it possible to reduce reliance on retrospective reports as to what *S* does during a filled interval prior to recall.

The results of Exp. IV indicate that the locus of forgetting is not in the transformed material, but rather that the transformed material is uniformly better recalled than would be predicted from its serial position. The act of selecting material for transformation presumably serves to focus attention on it and to isolate it (Burnstein, 1962). Forgetting occurs in material which has been presented just prior to the transform and decreases with the distance between transform and stored information.

Since it has been shown that information reduction and task difficulty, as measured by *Ss*' responses, are closely related for the particular transforms used in this study the results discussed so far stand independently of this particular a priori measure. That is, a response defined measure of difficulty of transform could be substituted for information reduction without changing any of the conclusions. On the other hand, there are a large number of tasks beside those employed here which can be said to involve information reduction. If it proves possible to specify the difficulty of such transforms in terms of the amount of information reduced the question of the relation between memory and information processing might be handled in a more general way.

One must be careful, however, in seeking to generalize the relation between information reduction and task difficulty beyond those tasks which have been specifically studied. Clearly, one can devise tasks with the same amount of information reduction which differ widely in difficulty. The degree of familiarity of the operations involved in the transformations will be a most important factor. Nevertheless, Pierce and Karlin (1957) have shown that it required more time to classify names into male and female than to call them out despite the great familiarity of both tasks. It takes longer to classify a two-digit number into odd and even than to name it (Posner, 1962), even though the classification task requires processing of only the final digit. It seems doubtful that combining numbers such as in addition ever becomes as fast as naming them regardless of the extent of training. It is not likely, therefore, that difficulty is only a matter of practice. It cannot be argued that the component operations of the subset of tasks used in this study are equal in familiarity although they were specifically chosen to involve highly familiar components. Modest amounts of training (Posner,

1962) did have more affect upon performance in the classification tasks than upon those involving addition although no change in the fundamental relations between information reduction and difficulty was found. Since even the relatively well-established relationship between difficulty and information transmitted in standard reaction-time tasks varies according to the familiarity of the stimulus-response relations (compatibility), it should be no surprise that similar considerations are important in reduction tasks as Fitts and Biederman (1965) have recently shown.

Because it is very difficult to equate tasks for familiarity when studying utilization of already well-learned transformation rules, much of the effort to determine the generality of the relation between information reduction and performance must be done by studies of the learning of new transformations. Posner (1964c) has shown that the amount of uncertainty between two patterns to be given a single name predicts the rate at which this classification can be learned. Moreover, a review of the literature in the area of concept learning (Posner, 1964b) has shown that the difficulty of concept attainment appears to be a direct function of the total amount of information within a category. These findings lend some support to the utility of the information-reduction measure in the study of information transformations.

The result of the present study showing that forgetting is systematically related to the difficulty of the information processing which follows the presentation of a stimulus is of importance for the study of many complex sequential tasks. Performance in such tasks as concept learning (Hunt, 1961), problem solving (Simon & Kotovsky, 1963), and serial arithmetic (Sampson, 1958) has been shown to be closely related to their storage requirements. Since many of these tasks may be described in terms of information transformations it might be possible to examine the interaction of memory and thought in these situations in a quantitative way.

REFERENCES

BROADBENT, D. E. *Perception and communication.* New York: Pergamon, 1958.

BROWN, J. Some tests of the decay theory of immediate memory. *Quart. J. exp. Psychol.*, 1958, **10**, 12–24.

BURNSTEIN, E. Some effects of cognitive selection processes on learning and memory. *Psychol. Monogr.*, 1962, **76**(35, Whole No. 554).

FITTS, P. M., & BIEDERMAN, I. S-R compatibility and information reduction. *J. exp. Psychol.*, 1965, **69**, 408–412.

HUNT, E. B. Memory effects in concept learning. *J. exp. Psychol.*, 1961, **62**, 598–604.

MCGEOCH, J. A., & IRION, A. L. *The psychology of human learning.* New York: Longmans, Green, 1951.

MELTON, A. W. Implications of short-term memory for a general theory of memory. *J. verbal Learn. verbal Behav.*, 1963, **2**, 1–21.

PETERSON, L. R., & PETERSON, M. Short-term retention of individual items. *J. exp. Psychol.*, 1959, **58**, 193–198.

PIERCE, J. R., & KARLIN, J. E. Reading rates and the information rate of a human channel. *Bell Sys. tech. J.*, 1957, **36**, 497–516.

PILLSBURY, W. B., & SYLVESTER, A. Retroactive and proactive inhibition in immediate memory. *J. exp. Psychol.*, 1940, **27**, 532–545.

POSNER, M. I. An informational approach to thinking. *USAF OSR Rep.*, No. 2635, 1962 (ASTIA No. 276136).

POSNER, M. I. Immediate memory in sequential tasks. *Psychol. Bull.*, 1963, **60**, 333–349.

POSNER, M. I. Effect of rate of presentation and order of recall in immediate memory. *Brit. J. Psychol.*, 1964, **55**, 303–306. (a)

POSNER, M. I. Information reduction in the analysis of sequential tasks. *Psychol. Rev.*, 1964, **71**, 491–513. (b)

POSNER, M. I. Uncertainty as a predictor of similarity in the study of generalization. *J. exp. Psychol.*, 1964, **68**, 113–118. (c)

SAMPSON, H. Stimulus duration and paced performance. *Canad. J. Psychol.*, 1958, **12**, 7–12.

SIMON, H. A., & KOTOVSKY, K. Human acquisition of concepts for sequential patterns. *Psychol. Rev.*, 1963, **70**, 534–546.

(Received July 7, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 5, 506-509

TRANSFER OF A PATTERN VERSUS COMPONENT DISCRIMINATION FOLLOWING TRAINING IN A PROBABILISTIC SITUATION[1]

MARCIA D. JOHNS

Colorado College

48 Ss were required to discriminate between stimulus patterns containing common components in a 2-choice probabilistic, discrimination-training situation. Training was followed by a series of test trials during which new stimulus combinations of the training components were presented. Discrimination during training was better described by a pattern interpretation than by a component interpretation of stimulus-sampling theory. Transfer of responding from training to test combinations occurred: (a) relative to a weighting of the separate components, or, alternatively (b) relative to the separate components, unless training combinations appeared as a part of the test display. Under the latter condition a more conspicuous role was given to the pattern interpretation of transfer behavior.

This study provides a test of pattern vs. component interpretations by means of a series of tests containing new combinations of stimulus components which appear during a preliminary series of probabilistic discrimination training trials. This study is an extension of the Estes and Hopkins (1961) paper which also provided such a transfer test. In the Estes and Hopkins study, however, the nature of the discrimination training (the classical type, rather than the probabilistic type) was such that the pattern interpretation could not be given a sharp test.

A modified component interpretation (presented by Atkinson, Calfee, Sommer, Jeffrey, & Shoemaker, 1964) is also tested in this study. (The modified interpretation did not appear in the Estes and Hopkins paper.)

Typically, on each trial in a two-choice discrimination (whether classical or probabilistic) S is presented with a combination of stimulus components, and is required to make one of two discriminative responses, called A_1 and A_2, which represent predictions that reinforcing event E_1 or E_2, respectively, will terminate the trial. The type of trial on which E_1 occurs is called a T_1 trial, and on which E_2 occurs, a T_2 trial.

In a classical situation, some stimulus components (unique, or relevant cues) are available for sampling only on T_1 trials or only on T_2 trials, while other components (common, or irrelevant cues) may be sampled on both T_1 and T_2 trials. In a probabilistic procedure (described by Burke & Estes, 1957), all components may be sampled on both T_1 and T_2 trials. The stimulus conditions associated with the two types of trials are differentiated only with respect to the distribution of probabilities with which the components are sampled. This differentiation provides the only basis for discrimination learning.

---

[1] A portion of the present paper is based on a dissertation submitted to the Graduate College of Indiana University in partial fulfillment of the requirements for the PhD degree. The writer is indebted to W. K. Estes for his guidance during the course of this research, and to G. R. Johns for his help and suggestions. The preparation of this paper was supported by funds from the National Science Foundation Grant GB-1725.

METHOD

The apparatus and stimulus materials were identical to those of Estes and Hopkins. There were four *S*'s booths each containing a vertical panel on which were mounted two lights side by side and a lever which could be moved to the left or right. The stimuli were four letters from the Russian alphabet which were projected on a screen that was visible from each of the booths.

The instructions (an expansion of the Estes-Hopkins instructions) described a medical analogue to the task of probabilistic discrimination learning. The *S* was asked to imagine a series of records of patients each of whom had one of two diseases (indicated by the onset of the left or right light at the end of a trial), to be diagnosed by *S* (by moving the lever to the left or right) from various combinations of four symptoms (represented by the Russian letters projected on the screen). The two directions of lever movement and, correspondingly, the two light locations were assigned in a counterbalanced manner to the roles of A_1 and A_2 and to the roles of E_1 and E_2, respectively.

Forty-eight *S*s were given, for the first part of the procedure, two sessions of 300 training trials each on two concurrent pattern vs. component discriminations of the probabilistic type, as shown in Table 1. The sampling probabilities for the components are conditional on the type of trial prevailing, T_1 or T_2, which were equally likely. The sampling-probability values were chosen in order to differentiate between the pattern and component outcomes which are discussed later, and to produce equal frequencies of presentation of the seven possible training patterns: *a.* (component *a* and not component *b*), *.b*, *ab*, *c.*, *.d*, *cd*, and "blank" (no components). During training, two concurrent discrimination problems, occurring equally often, were given so that only the above training patterns could be presented. Thus, the remaining possible combinations of the four components were reserved for presentation as test compounds. Presentations of the test compounds were interspersed among the last 50 training trials of both sessions. Each test compound was presented twice during a daily session, and on each test trial neither E_1 nor E_2 occurred.

TABLE 1

SAMPLING PROBABILITIES FOR THE COMPONENTS OF TWO CONCURRENT DISCRIMINATION PROBLEMS

Component	T_1	T_2
Problem I		
a	1	1/7
b	10/21	2/3
Problem II		
c	10/21	2/3
d	1	1/7

TABLE 2

OBSERVED AND PREDICTED PROPORTION OF A_1 RESPONSES TO TRAINING PATTERNS ON THE LAST 50 TRIALS OF DAY 2

Training Pattern	Observed	Predicted	
		Component	Pattern
a., *.d*	.97	.88	.92
.b, *c.*	.02	.42	.00
ab, *cd*	.94	.64	.83
"blank"	.01		.00

On all training trials the stimulus pattern was projected on a screen for 2 sec.; there was an empty interval of 1 sec.; one of two reinforcing lights was illuminated for 1 sec.; then, there was another empty interval of 2 sec. before the next pattern was shown. On all test trials the usual 1-sec. reinforcement interval was empty.

TABLE 3

OBSERVED AND PREDICTED PROPORTION OF A_1 RESPONSES TO THE TEST COMPOUNDS ON DAY 2

Test Compound	Obs.	Predicted			
		Patt.	Comp.	Mixed	Mod. Comp.
bc	.49	.02	.48	.48	.48
ac, *bd*	.87	.49	.72	.72	.77
ad	.96	.97	.95	.95	.95
abc, *bcd*	.71	.48	.64	.71	.69
abd, *acd*	.85	.95	.79	.95	.84
abcd	.84	.94	.72	.94	.77
Mean Absolute Diff. Bet. Obs. and Predicted		.22	.07	.05	.04

RESULTS

Tables 2 and 3 present predicted and observed response proportions to the training patterns and to the test compounds, respectively. From Table 1 and from the theoretical rules given by Estes and Hopkins, one may calculate the predictions from the component and pattern interpretations that appear in Tables 2 and 3.

Consider, e.g., the predictions of the probability of an A_1 response in the presence of training pattern *a.* (or *.d*). According to the pattern interpretation, responding is said to approach the probability of E_1 occurring in the presence of that pattern. The latter probability is .92, shown in Table 2 and calculated from the appropriate values in Table 1. Under the component interpretation, on the other hand, responding to pattern *a.* is said to approach the probability of E_1 occurring in the presence of component *a*, whether alone or in combination with component *b*. This E_1 probability, calculated from Table 1, is .88. The component prediction for responding to the two-letter pattern *ab* is .64 which is, according to a compounding rule (Burke & Estes, 1957), the average of the separate component predictions to patterns *a.* (.88) and *.b* (.42).

For the training data in Table 2, the pattern interpretation was clearly more satisfactory than the component interpretation, as in the Estes-Hopkins study. However, the training data were not thoroughly compatible with the pattern interpretation due, in part, to the observed tendency of some *S*s to maximize their chances for success on each trial.

Since neither of the interpretations for the training data was uniformly satisfactory, transfer predictions of the test data, shown in Table 3, were based on the observed (rather than predicted) response proportions to stimulus units achieved during training, shown in Table 2.

Consider the application of the compounding rule to obtain predictions for responding to test compound *abc*, e.g. The pattern prediction, .48, was obtained by averaging .94 (the observed value to pattern *ab*, Table 2) and .02 (pattern *c.*). The component prediction, .64, was obtained by averaging .95 (the observed value to component *a*, calculated from Table 2), .48 (component *b*), and .48 (component *c*).

For the test data in Table 3, the pattern interpretation (called the "molar" pattern interpretation by Estes and Hopkins) was clearly the poorest, while the component interpretation was considerably better, as indicated by .22 and .07, respectively—the mean absolute difference between observed and predicted response proportions. On the other hand, Estes and Hopkins reported the superiority of the pattern interpretation over the component interpretation of the test data. However, for classical discrimination training (as used by Estes and Hopkins) the pattern interpretation cannot be differentiated from the component interpretation when the latter interpretation is extended (as suggested by Johns, 1962) by an auxiliary process such as the acquisition of an observing response (Wyckoff, 1952) or the adaptation of irrelevant cues (Restle, 1955). Under such an extended component interpretation *S* may be responding to the constituents of the stimulus combinations and, at the same time, may be learning to make the correct response to, say, *ab* by ignoring component *b* when it appears with component *a*. (For a probabilistic procedure, as used in the present study, there is no obvious reinforcement of *Ss'* learning to make observing responses or to adapt to certain cues.)

A mixed interpretation, which ranks second best in Table 3, combines elements of both the pattern and component interpretations. The compounding rule for the mixed case requires that whenever the test compound contains a training pair, then the response value to that pair is averaged with the response value to the remaining component. The rule produces identical predictions for the component and mixed interpretations of all test pairs (since a test pair cannot be a training pair), and identical predictions for the pattern and mixed interpretations of the combination of all four components.

It should be noted that the best interpretation for the test data in Table 3 is a modified component interpretation which also provides the best fit for data on stimulus compounding recently obtained by Atkinson et al., 1964. This interpretation (called Model II by Atkinson et al.) assumes that the influence of each component is proportional to the likelihood that it elicits a correct response. Thus, for the test compound *abc*, e.g., the modified component prediction, .69, was obtained by weighting each proportion of A_1 responses—.95 (component *a*), .48 (*b*), and .48 (*c*)—by the appropriate observed proportion of correct responses—.84 (component *a*), .50 (*b*), and .50 (*c*), and dividing by the sum of the proportions correct, 1.84. The superiority of the modified component interpretation in the present study, as shown in Table 3 by a difference of .04 between observations and predictions, is not, however, striking since the mixed interpretation ranks closely to it with a value of .05.

REFERENCES

ATKINSON, R. C., CALFEE, R. C., SOMMER, G. R., JEFFREY, W. E., & SHOEMAKER, R. A test of three models for stimulus compounding with children. *J. exp. Psychol.*, 1964, 67, 52–58.

BURKE, C. J., & ESTES, W. K. A component model for stimulus variables in discrimination learning. *Psychometrika*, 1957, 22, 133–145.

ESTES, W. K., & HOPKINS, B. L. Acquisition and transfer in pattern versus component discrimination learning. *J. exp. Psychol.*, 1961, 61, 322–328.

JOHNS, M. D. *Transfer of a pattern versus component discrimination following training in a probabilistic situation.* (Doctoral dissertation, Indiana University) Ann Arbor, Mich.: University Microfilms, 1962, No. 63–3833.

RESTLE, F. A theory of discrimination learning. *Psychol. Rev.*, 1955, 62, 11–19.

WYCKOFF, L. B., JR. The role of observing responses in discrimination learning. *Psychol. Rev.*, 1952, 59, 431–442.

(Received July 8, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 5, 510-519

ADJUSTMENT TO NEW REWARD: SIMULTANEOUS- AND SUCCESSIVE-CONTRAST EFFECTS

NORMAN E. SPEAR AND WINFRED F. HILL

Rutgers University *Northwestern University*

In 2 experiments, rats choosing between 2 rewards in a position discrimination experienced a change in either the quantity or quality of reward offered on the more favorable alternative (MFA) while reward on the less favorable alternative (LFA) was held constant. Experience with the LFA and MFA was equated throughout by interspersing forced trials among free choices. Before the reward shift, (a) maximal choice of the MFA was attained more quickly the greater the reward there, and (b) running speed to the LFA was inversely related, and more sensitive than MFA speed, to magnitude of reward on the MFA. After the reward shift, (a) all response measures eventually adjusted to the level of the base line control, though speed on the LFA did not adjust as readily as MFA speed, and (b) some tendency appeared for speed on the MFA to initially undershoot (not statistically significant) the appropriate level when reward was drastically reduced there.

Change in the magnitude of reinforcement, manipulated by altering nominally or operationally defined reward, has been reflected by rapid and marked changes in behavior. The widely cited work of Crespi (1942) has been followed by a number of studies of a similar nature (e.g., Collier & Marx, 1959). The basic approach has been to either raise or lower the (nominal) reward value which initially had been given *S*. Thus, from the first to the second successive, operationally discrete stage of training, treatment in this paradigm ("successive contrast") differs only in terms of contrasting magnitudes of reward.

A second, more recent, set of studies has presented different magnitudes of reward simultaneously. In these experiments, *S*'s preference in terms of a choice measure has been measured along with independent measures of *S*'s speed of approach to each stimulus when presented singularly (e.g., Spear, 1964). This paradigm may be termed a "simultaneous contrast" in reward.

The present experiments incorporated both of these paradigms within a single design. Reward on the nominally less favorable alternative (LFA) of a position discrimination was held constant across groups while varying reward on the more favorable alternative (MFA); this represented the simultaneous-contrast paradigm. After a moderate amount of training under these conditions, reward on the MFA was changed to a common level for all groups (less than before for some groups; more than before for others), with no change on the LFA; this represented the successive-contrast paradigm. This design raises a number of questions concerning differential speed between the alternatives within groups and differential speed to a given alternative between groups, along with the choice measure.

In this design a simultaneous-contrast effect would be said to occur if performance on the LFA were inversely proportional to the nominal

---

[1] This research was supported by National Science Foundation Grant GB-254, United States Public Health. MH-08770-01 and MH-08888-01, and Rutgers Research Council Grant No. 686.

amount of reward on the MFA. A successive-contrast effect would occur if performance on the MFA after the reward shift were inversely proportional to the reward that had been given there before the shift.

EXPERIMENT I

Method

Subjects and apparatus.—The *Ss* were 84 naive, male albino rats of the Sprague-Dawley strain, 98–107 days old at the start of T-maze experience. Three additional *Ss* were discarded for failure to enter an arm of the maze within 2 min. after leaving the start box for six successive trials. The maze was the narrow T (14-ft. stem and 9-in. arms, excluding goal box) previously described by Cotton, Lewis, and Metzger (1958). The formerly opaque forcing doors were replaced with sheet-metal cloth; this permitted visual and olfactory access to both alternatives during forced trials. Stem speed was measured from the treadle just outside the start box to the treadle just before the choice point. Arm speed was measured from the second treadle to either of the third treadles which were positioned just before each goal box (distance of 12 in.).

Procedure.—Initial taming proceeded as described by Spear (1964), and the feeding schedule consisted of 1¼ hr. of free feeding, which period began 1¼ hr. after the start of the day's scheduled T-maze session. Thus, *Ss* were always run under 21½-hr. food deprivation.

There were six groups to which *Ss* were assigned at random in this experiment, and three members from each group (i.e., 18 *Ss* in all) were run in rotation within a 1¼-hr. period. Each *S* was given six trials per day for 16 consecutive days. A minimum of 10 min. occurred between successive trials. The *S* was detained between trials in an individual carrying-cage compartment where water was always available.

Upon entry into a goal box, *S* was detained for a minimum of 15 sec., or until all pellets were consumed. The first and fifth trials of each day were free choices. On the second and sixth trials, *S* was forced to the alternative opposite that chosen on the preceding free trial. On one of the third and fourth trials, *S* was forced to the MFA, and on the other *S* was forced to the LFA. The direction of forcing on the third and fourth trials was determined according to one of two previously set, random schedules. These schedules were constructed and assigned so that, on any given day, half the *Ss* in each group went to the MFA (which was the left arm for a random half of the *Ss*, and the right arm for the remaining *Ss*) on Trial 3 of that day, and half went to the MFA on Trial 4. Unless otherwise specified, arm-speed data are based upon speeds on these randomly forced trials.

Design.—Differential treatment was introduced in terms of the magnitude or sweetness of the reward present in the MFA. In the case of the group given equivalent reward in both alternatives, the MFA was randomly designated prior to training. All *Ss* were always rewarded with 1 regular pellet in the LFA, and either 1, 2, or 10 pellets in the MFA. Pellets in the MFA were regular for three groups and sucrose (also Noyes pellets, of the same weight and dimension as the regular pellets) for the other three groups. The groups may be designated according to their conditions as 1Sw, 2Sw, or 10Sw (for sweet), and 1R, 2R, or 10R (for regular). This differential treatment was given during only the first 8 days of T-maze experience—the Preshift stage of training. From the ninth through sixteenth days, all *Ss* were always presented with 2 regular pellets in the MFA and their usual 1 regular pellet in the LFA; this was termed the Postshift stage of training. Therefore, simultaneous-contrast effects could be defined in terms of group differences in speed to the LFA during the Preshift stage. Successive-contrast effects could be defined during the Postshift stage if a group's mean speed to the MFA should either overshoot or undershoot the corresponding mean speed of Group 2R.

Results

Choices.—Eventually, *S*s from all groups (except those choosing between equivalent rewards) unanimously chose the MFA on their free trials, though the number of trials necessary to reach this asymptotic choice level depended upon the reward present in the MFA. The course of choice behavior throughout both stages of training is shown in Fig. 1.

During the Preshift stage, both the effects of sweetness, $F(1, 78) = 4.16$, $p < .05$; and magnitude, $F(2, 78) = 16.01$, $p < .001$, were significant with no interaction, $F(2, 78) < 1$. The

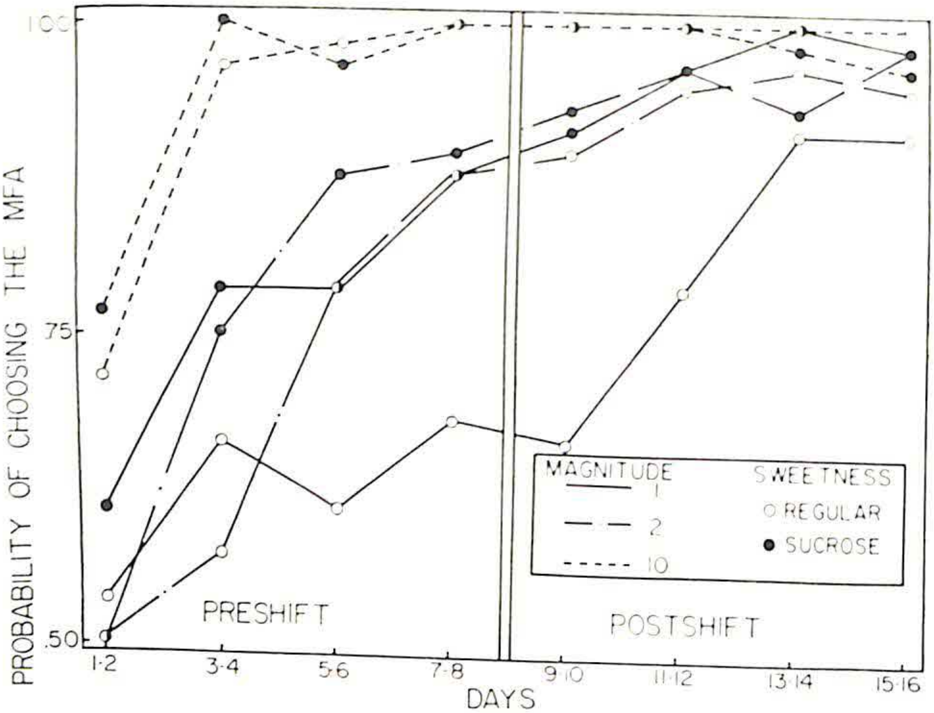


FIG. 1. Probability of choosing the MFA during the Preshift and Postshift stages of Exp. I. (The MFA for Group 1R was randomly determined prior to training, though reward was the same in both alternatives during Preshift.)

greater the sweetness or the greater the magnitude, the greater was the (arc sine) probability of a correct choice throughout the entire 8 days. The effect of sweetness, however, was of only borderline significance. When multiple comparisons were made via the Duncan range test (Edwards, 1960), Groups 10R and 10Sw differed from all the others (though not from each other), but the remaining four groups did not differ significantly among themselves.

It is clear from Fig. 1 that all groups continued to choose the MFA during the Postshift stage. Although Group 1R is somewhat below the others during the latter portion of this stage, 8 of the 10 errors made by this group over the last 4 days were made by the same *S*.

Arm speed during preshift stage.—The basic results may be summarized as: (*a*) slower running to the LFA the greater the reward in the MFA; and (*b*) no significant differences in speed of running to the MFA. These relationships may be seen in Fig. 2.

An analysis of variance of mean speed to the LFA during the last 2 days of the Preshift stage was performed. The apparent inverse relationship between speed to the LFA and reward on the MFA was verified in terms of the main effect of nominal magnitude, F (2, 78) = 8.85, $p < .001$. Neither the main effect of sweetness, F (1, 78) < 1, nor the interaction between sweetness and magnitude, F (2,78) = 1.32, attained significance at the .05 level.

Speed to the MFA during the final 2 days of Preshift training did not vary significantly as a function of magnitude, F (2, 78) = 1.95, $p < .25$, nor sweetness, F (1, 78) < 1, nor did these two variables interact significantly, F (2, 78) < 1. In view of the rapid ac-

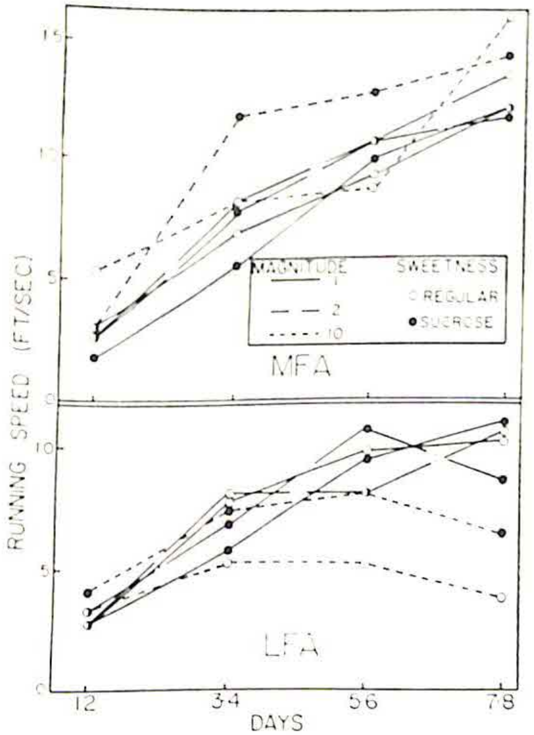


Fig. 2. Mean arm speed to the MFA (top) and to the LFA (bottom) during the Preshift stage of Exp. 1.

celeration of Group 10Sw in terms of speed to the MFA, a post hoc analysis of variance of MFA speed summed over all of Preshift training was also performed. Again no effect was found at the .05 level of magnitude, F (2, 78) = 2.63, $p < .10$; sweetness, F (1, 78) < 1; nor their interaction, F (2, 78) < 1. Thus, no statistically significant effect of reward magnitude on running speed to the MFA could be determined at any point during the Preshift stage. Moreover, there does not appear to be any suggestion from Fig. 2 that such differences might have appeared had additional training been given under these conditions.

Arm speed to the MFA on free trials yielded the same general relationships as were found on the forced trials, though it is believed that the forced-trial data is more meaningful, especially during the Preshift stage. Clearly, since all *S*s did not choose the MFA on any given trial, some undesirable *S* selection must occur in an analysis of free-trial arm speeds.

Analysis of stem speeds yielded no significant differences in either stage of training.

Arm speed during postshift stage.—The two major findings concerning this measure—(*a*) the absence of successive-contrast effects in the MFA; and (*b*) the paradoxically greater change of speed on the MFA than the LFA, which change depended upon original reward magnitude—are evolved below. The course of these running speeds with training are shown in Fig. 3. Parenthetically, a note should be injected concerning the spurious differences in speed to the MFA on the first 2 days of the Postshift stage. An analysis of variance of this measure revealed (as has been shown for the Preshift stage) no statistically significant effects of magnitude, F (2, 78) = 2.32, $p < .25$;

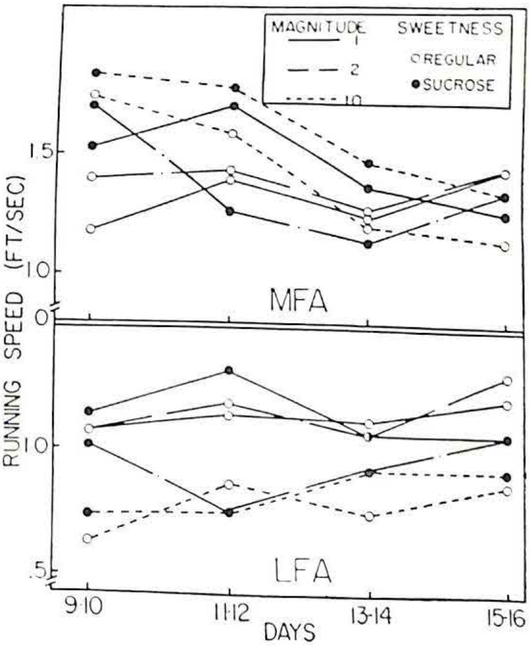


Fig. 3. Mean arm speed to the MFA (top) and to the LFA (bottom) during the Postshift stage of Exp. 1.

sweetness, $F\ (1, 78) = 2.37$, $p < .25$; nor of their interaction, $F\ (2, 78) < 1$.

As a test for successive-contrast effects, an analysis of variance was performed on mean running speeds to the MFA during the final 2 days of the Postshift stage. Neither the main effects of sweetness and magnitude, nor their interaction proved to be significant (all *F*s were less than 1). Thus, there was no evidence for successive-contrast effects as measured by absolute levels of speed to the MFA.

There is a second, perhaps more important, aspect to these Postshift data. This is the relative change in speed to the original MFA, and to the MFA following removal of differential conditions of nominal reward. In order to evaluate this, mean speed on combined Days 1 and 2 was compared with that on Days 7 and 8 of the Postshift stage. Thus, the basic analysis was a 3 × 2 × 2 analysis of variance comparing the effects of magnitude, sweetness, and the within-*S* measure of Days. This analysis indicated that speed to the LFA was unaffected by the change in reward magnitude to the MFA. The effect of magnitude persisted throughout the Postshift stage, $F\ (2, 78) = 4.40$, $p < .025$, such that the greater the Preshift magnitude on the MFA, the less the speed to the LFA. This speed measure did not differ as a function of Postshift experience—for Days effect: $F\ (1, 78) = 3.10$, $p < .10$—nor did the effect of Preshift magnitude change with increased training on the Postshift magnitude—for Days × Magnitude interaction, $F\ (2, 78) < 1$. There was no significant effect of sweetness, $F\ (1, 78) < 1$; no significant interaction of Sweetness × Days, $F\ (1, 78) = 1.34$, $p < .50$; and all remaining interactions from this analysis yielded *F*s of less than 1.

A 3 × 2 × 2 analysis of variance identical to that above was also performed on speed to the MFA. As may be inferred from Fig. 3, the effect of the shift in reward was much greater on the MFA than was the case with speed measured in the LFA. The overall change in speed from Days 1 and 2 to Days 7 and 8 was substantial, $F\ (1, 78) = 8.17$, $p < .01$. Moreover, this change depended upon Preshift magnitude—for the interaction between Magnitude × Days: $F\ (2, 78) = 4.09$, $p < .025$—such that the greater the shift in reward magnitude, the more the speed change from the beginning to the end of the Postshift stage. No other sources of variance from this analysis were statistically significant.

It is clear, then, that the reward shift had different effects on the MFA than those on the LFA. Specifically, the magnitude of the change in speed to the MFA was greater the greater the shift in reward; but speed to the LFA did not change, regardless of the magnitude of the MFA reward shift. Paradoxically, speed to the LFA had been the more sensitive to absolute reward magnitude on the MFA prior to the shift.

Discussion

While no successive-contrast effect appeared, the simultaneous-contrast effect was clear and reliable in terms of the arm-speed measure. But the asymmetry of the effect of reward magnitude—slower LFA speed was not accompanied by correspondingly faster MFA speed in groups with largest MFA reward—remained puzzling. Although preference for the MFA in terms of choices was clearly attained more quickly the greater the MFA reward (Davenport, 1962; Hill & Spear, 1963; Spear, 1964) the differential in speeds to the alternatives was notably attributable to behavior on the LFA only.

It is true that this lack of speed differences on the MFA agreed with the results of Davenport (1962) from an identical (Preshift) paradigm, and with the results of others (e.g., Goldstein & Spence, 1963; Spear, 1964) who have used similar paradigms. Nevertheless, the fact was difficult to rationalize, and it was felt that a closer look at the response components comprising the arm-speed measure was warranted. Therefore, a second experiment gave independent consideration to the "turning" and the "committed" speeds—the respective speeds with which *S* left the stem and completely entered an arm, and the final speed to the goal box once *S* was directly in line with the food cup.

Parenthetically, it should be explained that the search for contrast effects with bulk of reward relatively constant—the use of sucrose vs. regular pellets—was dropped for Exp. II. This variable had produced negligible differences in Exp. I with the exception of the tendency for Group 1Sw to choose a sucrose over a regular pellet on free choices. Thus, Exp. I and other evidence (e.g., Goodrich & Zaretsky, 1962) suggests that the strong contrast effects shown by Collier and Marx (1959) were not due solely to their varying quality (percentage of sucrose) rather than bulk of reward.

EXPERIMENT II

This experiment compared Group 1R with a group choosing between two diverse magnitudes (12 vs. 1 pellets; Group 12R) in order to accomplish the following:

1. The phenomena associated with the speed measure employed in Exp. I were analyzed in further detail by independent consideration of two components of this measure—the "turning speed" and the "committed speed." These phenomena were also examined in Exp. II for their persistency, with particular attention given to the endurance of the simultaneous-contrast effect—the differential effect of MFA reward magnitude on speed to the LFA.

2. The possibility of choice changes was increased by modifying the design. An original purpose of Exp. I had been to examine contrast effects measured in terms of choice behavior. But the base line control, Group 2R, continued to increase toward 100% choice of the MFA during the Postshift stage. Thus, the lack of preference change (as measured by choice) by Groups 10Sw and 10R is not surprising since their absolute level of differential speed never dropped significantly below that of Group 2R. That is to say, this fact is predictable provided one assumes that the absolute level of differential speed dictates choice to a greater extent than the amount of *change* in differential speed. In any case, ceiling effects such as 100% choice could conceivably have masked the effects of the reward shift on choice behavior. Such limiting effects on this measure were eliminated in Exp. II by reducing the magnitude of reward on the MFA to the same level as that on the LFA. Also, Postshift training was extended to provide greater opportunity for change in choice behavior.

Method

Subjects and apparatus.—Two groups of 16 female albino (Sprague-Dawley) rats, 75–85 days old, were run in a T maze which differed from that used in Exp. I in that it consisted of a 10-in. start box, 24-in. stem, 18-in. arms, and 12-in. goal boxes; all parts were 4 in. wide and 4 in. high. Three measures of running time (later converted to speeds) included: "stem" (time elapsed from the opening of the start-box door until *S* interrupted a beam located 18 in. down the stem), "turning" (beginning 6 in. before the choice point and ending 6 in. beyond in either arm), and "committed" (beginning 6 in. beyond the choice point in either arm—at which point *S* could not retrace to the stem or remaining alternative—and continuing for

18 in. down either arm to a point 2 in. before the pellet container). As in Exp. I, the speeds referred to are those recorded on the randomly determined forced trials.

Procedure.—Prior to T-maze experience, *S*s were treated as in Exp. I except that (*a*) taming was given without food pellets available, and (*b*) *S*'s daily ration was 9 gm. of chow and a total of 30 regular .045-gm. Noyes pellets, which total included reward consumed in the T maze.

The T-maze training proceeded as in Exp. I except that only four trials were given each day. For a randomly selected half of the *S*s in each group the first trial was a free choice, on the second trial *S* was forced to the side opposite that just chosen, and the third and fourth trials were "randomly forced" (one to each alternative). The remaining *S*s received their randomly forced trials *prior* to their free trial each day. The particular forcing procedure included the characteristic that, on any given trial throughout the experiment, one fourth of the *S*s in each group were forced to their MFA (one eighth of the *S*s to the left and one eighth to the right), and one fourth of the *S*s in each group were forced to their LFA.

For the first 48 trials, *S*s in Group 12R received 12 pellets upon each experience in the MFA and 1 pellet in the LFA, and *S*s in Group 1R received 1 pellet whichever alternative they experienced. During the 72 trials of the Postshift stage, the groups were treated identically and received 1 pellet in both alternatives. Thus, treatment in the Preshift and Postshift stages was the same except for the reduction in MFA magnitude available to Group 12R.

Results

Choices.—The course of choice behavior in the Preshift and Postshift stages is shown in Fig. 4. By the end of acquisition, as expected, *S*s in Group 12R showed maximal choice of the MFA and *S*s in Group 1R were choosing it about half the time.

After the shift in MFA magnitude of reward, choice of the MFA gradually declined in Group 12R. After 72 trials of Postshift experience, *S*s in this group were choosing the original MFA with about the same frequency as the remaining alternative. Clearly, a "depression effect" in choice behavior, definable by fewer choices of the original MFA by Group 12R compared with Group 1R, did not occur.

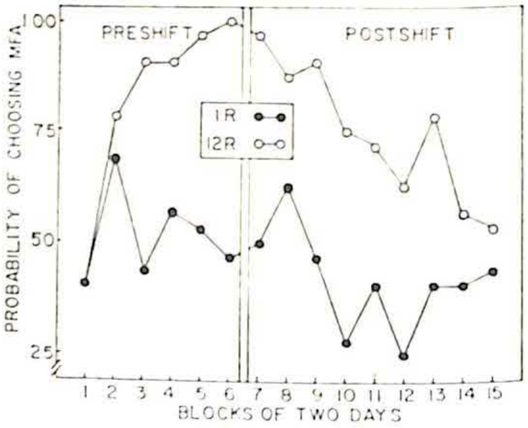


FIG. 4. Probability of choosing the MFA during the Preshift and Postshift stages in Exp. II.

Turning and committed speed during the preshift stage.—It is first important to establish basic replication of the running-speed phenomena found in Exp. I. The arm-speed measure and conditions of Exp. I would be most clearly replicated by considering the speed for the combined turning and committed distances for those *S*s whose randomly forced trials were given as the third and fourth trials of each day (i.e., after each alternative had already been sampled once in a given session, as was the case in Exp. I). With this arm-speed measure, the results from Exp. I were completely replicated in Exp. II. Speed to the LFA by Group 12R was less than that by Group 1R, t (14) $= 2.56$, $p < .05$; and the greater speed to the MFA by Group 12R again did not attain statistical significance, t (14) $= 1.91$, $p > .05$.

The question now is the way in which this effect is distributed among the two distinguishable components of the response chain—the turning speed and the committed speed. These component speeds for the Preshift stage are shown in Fig. 5.

Analyses on the last 2 days of Preshift revealed that turning speed for Group 12R was both greater to the MFA, t (30) = 4.15, $p < .001$, and less to the LFA, t (30) = 2.71, $p < .02$, than that for Group 1R. Thus, a simultaneous-contrast effect was found on the LFA with this measure; and, in addition, speed was also greater to the MFA the greater the reward there. This latter fact was not true in terms of committed speed, which accounts for the lack of a significant difference when these measures are combined. That is, committed speed to the MFA did not differ as a function of reward magnitude there, t (30) = .01, $p > .90$. Committed speed to the LFA at the end of the Preshift stage was not quite significantly greater for Group 1R than Group 12R, t (30) = 1.55, $p > .10$, but the greater LFA speed by Group 1R recorded on the first 2 days of the Postshift stage did attain statistical significance, t (30) = 2.12, $p < .05$, which was maintained through several days of the Postshift stage.

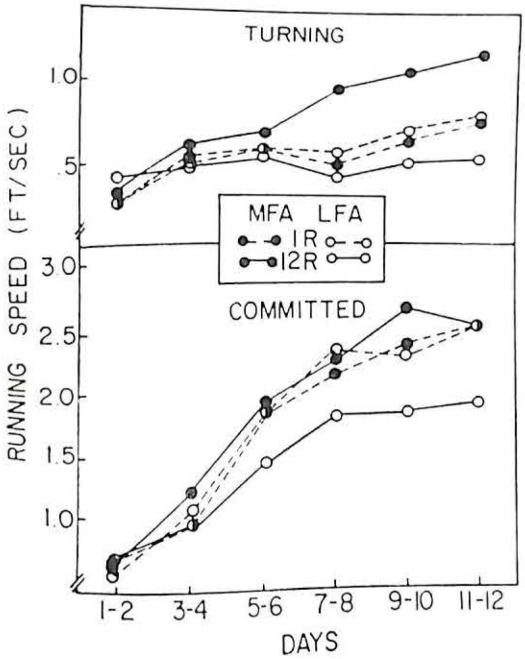


FIG. 5. Mean turning speed (top) and mean committed speed (bottom) to the MFA and LFA during the Preshift stage of Exp. II.

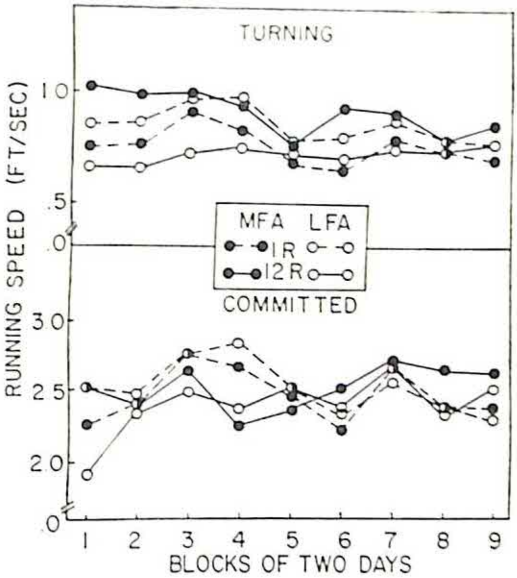


FIG. 6. Mean turning speed (top) and mean committed speed (bottom) to the MFA and LFA during the Postshift stage of Exp. II.

Stem speeds were nearly identical for the two groups, in both stages in the T_4^- maze.

Turning and committed speed during the postshift stage.—It is apparent in Fig. 6 that the measure most quickly and markedly affected by the shift in reward magnitude available on the MFA was speed to the MFA. This fact is in complete agreement with Exp. I. By the seventh and eighth days of this stage (the point, in terms of days, at which Postshift training in Exp. I was terminated), turning speed to the MFA did not differ between groups, t (30) = 1.03, $p > .30$, and the group relationship in terms of committed speed had reversed. This slower committed speed to the formerly MFA by Group 12R (a successive-contrast effect, by definition) did not attain statistical significance at this point, t (30) = 1.89, $.10 > p > .05$. As in Exp. I, an analysis of variance (on MFA committed speed) with groups as one factor and Postshift Days 1 and 2 vs. Days 7 and 8 as the other, did reveal a significant interaction, F (1, 30) = 13.01, $p < .005$, which indicates that the change in group differences was not a chance occurrence. Of course, this interaction is not sufficient in itself to define a successive-contrast effect.

In contrast to the speeds to the former MFA, the simultaneous-contrast effect in terms of speeds to the former LFA was quite persistent. On the seventh and eighth days of the Postshift stage, turning and committed speeds to the former LFA were still significantly greater for Group 1R than for Group 12R, t (30) = 2.17, $p < .025$; and t (30) = 2.37, $p < .025$, respectively.

By the end of the Postshift stage, all speed differences were gone. On the last 2 days of this stage, the groups' turning speeds did not differ to the MFA, t (30) = 1.30, $p > .20$, nor to the LFA, t (30) = .09, $p > .90$; neither did committed speed differ to the MFA, t (30) = 1.18, $p > .20$, nor to the LFA, t (30) = .85, $p > .40$.

In summary, committed speeds adjusted more quickly than turning speeds, including an immediate decrease in speeds to the former MFA and an eventual increase in speeds to the former LFA, relative to Group 1R. After extensive Postshift training, all speed measures (and choice behavior) adjusted to a level appropriate to the current reward.

Discussion

The major facts obtained in Exp. I were substantiated and extended in Exp. II. First, it is clear in terms of speed measures that a simultaneous-contrast effect may exist when a reliable successive-contrast effect cannot, and therefore, the operational distinctions between the two paradigms could aid in understanding contrast effects in general. This implies that such considerations as depletion level of S (see, e.g., Ehrenfreund & Badia, 1962) may not be the only features to be considered when searching for the elusive successive-contrast effect, but that the characteristic distinctions between simultaneous- and successive-contrast effects are also important. For example, in view of the basic temporal distinction (due to the fact that trials in tests of successive contrast have nearly always been given at the rate of one per day by necessity) one might predict that simultaneous-contrast effects would weaken with increased intertrial interval. Indeed, in an experiment conducted at Rutgers University, the first author has collected data which suggests that the simultaneous-contrast effect changes markedly when a 24-hr. interval elapses between trials. The relative importance of other characteristic differences between these two paradigms, e.g., the greater number of "comparisons" (shifts) possible with the simultaneous-contrast paradigm, remains an empirical question. On the other hand, it is quite possible that the existence of the simultaneous-contrast effect may have itself decreased the likelihood of the successive-contrast effect. It is conceivable that the exposure in the Preshift stage to two different rewards may have reduced or eliminated the kind of anchoring effects which, in accord with a perceptual interpretation of these phenomena (cf. Collier & Marx, 1959), may be responsible for successive-contrast effects.

Secondly, it is clear from both experiments, and for both response components of the arm-speed measure, that adjustment in speed following a shift in MFA reward is more rapid in the former MFA than in the former LFA. Moreover, this is true even though speed to the LFA, relative to MFA speed, is apparently *more* sensitive to magnitude differences on the MFA *before* the shift. Of course, one could hardly expect the former relationship to be otherwise without attributing more cognitive abilities to the rat than are normally supposed. In particular, faster adjustment to the former LFA would require that *S* retain, during its response to the LFA, some representation of the reward currently offered on the former MFA; and this representation would need to be independent of, or at least uncorrelated with, *S*'s response to the MFA. In any case, the fact that *S*'s speed to the LFA eventually did adjust to the level of the base line group after extended Postshift training substantiates the transitory nature of contrast effects reported by Gonzales, Gleitman, and Bitterman (1962).

With respect to the above relationships, some benefit occurred from independent consideration of the turning- and committed-speed components of the response from the choice point to either alternative. A major advantage was the discovery that MFA speed did depend upon MFA magnitude with the turning measure alone, though not with the committed speed alone nor when these measures were combined. It may be significant that the asymptotic committed speed attained in the last 2 days of the Preshift stage (about 2.5 ft/sec to MFA for Group 12R and to both sides for Group 1R) was very similar to that of the asymptotic stem speeds (2.44 and 2.31 ft/sec for Groups 1R and 12R, respectively) at that point. This suggests that *S*s were running the "committed" portion of the T maze subsequent to turning at about the same rate (with the exception of LFA speed by Group 12R) as they did in the stem after starting from a still position in the start box. If, then, turning speed is analogous to the starting speed in the runway, one would expect greater differences there as a function of reward magnitude in comparison with points nearer the goal (cf. Knarr & Collier, 1962).

REFERENCES

Collier, G., & Marx, M. H. Changes in performance as a function of shifts in the magnitude of reinforcement. *J. exp. Psychol.*, 1959, **57**, 305–309.

Cotton, J. W., Lewis, D. J., & Metzger, R. Running behavior as a function of apparatus and of restriction of goal box activity. *J. comp. physiol. Psychol.*, 1958, **51**, 336–341.

Crespi, L. P. Quantitative variation of incentive and performance in the white rat. *Amer. J. Psychol.*, 1942, **55**, 467–517.

Davenport, J. W. The interaction of magnitude and delay of reinforcement in spatial discrimination. *J. comp. physiol. Psychol.*, 1962, **55**, 267–273.

Edwards, A. L. *Experimental design in psychological research.* New York: Holt, Rinehart, & Winston, 1960.

Ehrenfreund, D., & Badia, P. Response strength as a function of drive level and pre- and postshift incentive magnitude. *J. exp. Psychol.*, 1962, **63**, 468–471.

Goldstein, H., & Spence, K. W. Performance in differential condition as a function of variation in magnitude of reward. *J. exp. Psychol.*, 1963, **65**, 86–93.

Gonzalez, R. C., Gleitman, H., & Bitterman, M. E. Some observations on the depression effect. *J. comp. physiol. Psychol.*, 1962, **55**, 578–581.

Goodrich, K. P., & Zaretsky, H. Running speed as a function of concentration of sucrose incentive during pretraining. *Psychol. Rep.*, 1962, **11**, 463–468.

Hill, W. F., & Spear, N. E. Choice between magnitudes of reward in a T maze. *J. comp. physiol. Psychol.*, 1963, **56**, 723–726.

Knarr, F. A., & Collier, G. Taste and consummatory activity in amount and gradient of reinforcement functions. *J. exp. Psychol.*, 1962, **63**, 579–588.

Spear, N. E. Choice between magnitude and percentage of reinforcement. *J. exp. Psychol.*, 1964, **68**, 44–52.

(Received July 8, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 5, 520-525

INTRASESSION DECREMENTS IN THE PERFORMANCE OF THE CLASSICALLY CONDITIONED EYELID REFLEX[1]

WILLARD N. RUNQUIST AND WILLIAM R. MUIR

University of Alberta

3 groups of 19 Ss were given 500 conditioning trials spaced over either 1, 5, or 10 days. The 5-day group showed a decrease in percentage of CRs within postasymptotic sessions, while other groups did not. All groups showed increase in CR latency within these sessions, but the increase was greater in the 5-day group. These changes occurred despite overall increase in percentage of CRs across blocks of 100 trials for all groups.

This study is concerned with performance decrements which appear in the classical conditioning of the eyeblink reflex when conditioning is extended for a large number of trials. Although Pavlov (1927, Lecture 14) reports that well established CRs tend to disappear with continued training, there has been virtually no systematic investigation of the phenomenon.

Most investigations of classical defense conditioning utilizing the eyeblink reflex have not carried out the conditioning process far enough beyond acquisition to have observed any tendency for the CR to diminish. However, a recent series of studies by Prokasy and his associates (Ebel & Prokasy, 1963; Krauss & Prokasy, 1963; Prokasy, Ebel, & Thompson, 1963; Prokasy & Whaley, 1961, 1963) have utilized procedures whereby several hundred conditioning trials have been given spaced over several sessions. In only one of these studies (Prokasy & Whaley, 1963) were performance decrements reported and, in this case, decrements within a session (65 trials) were observed for most *Ss* after about 200 trials. Thus, the evidence for inhibition with reinforcement in eyelid conditioning is not compelling.

Since the number of trials given per day might be expected to be an important factor, *Ss* in this experiment were given 500 conditioning trials either in 1 day, spaced over 5 days, or spaced over 10 days.

METHOD

Subjects and design.—The *Ss* were 57 volunteers from an introductory psychology class and were paid for their services. Assignment to one of three groups was made more or less randomly with the restriction, of course, that an *S* had to be free to serve in the required condition. Since this restriction resulted in nonrandom assignment to conditions, some care must be exercised in comparisons between groups. The *Ss* in Group I each received 500 trials on 1 day with a 10-min. rest between each block of 100 trials. The *Ss* in Group V each received 100 trials per day for 5 consecutive days. The *Ss* in Group X received 50 trials per day for 10 days with a 2-day rest between Days 5 and 6.

Apparatus.—The laboratory consisted of two separate rooms, one containing the apparatus for programing and presenting stimuli and for recording responses, the other containing two 3 × 6 ft. booths for *Ss*, arranged so that two *Ss* could be run simultaneously. The inside of each booth was painted flat white and was illuminated by a single 40-w. fluorescent lamp located 10 ft. above the floor. A 2-in. black cross was painted on the wall at about eye level and acted as a fixation point.

---

[1] Supported in part by Grant MH-04753 from National Institutes of Health. John Roshak aided in construction of the apparatus and collected some of the data. We would like to express our appreciation to W. F. Prokasy, Pennsylvania State University, who provided most of the latency analysis.

The recording apparatus consisted of the mechanical-electrical linkage system now in common use. The pickup unit was a Hunter headset connected to *S*'s eyelid by a false eyelash taped to the lid. The blink was amplified and recorded with a Brush Mark II ink writing oscillograph; paper speed was 125 mm/sec. The CS was a tone of 760 cps produced by a Smith Model 2425 code practice oscillator and presented over Telex Model 1A earphones. The UCS was a puff of compressed nitrogen under 80-mm. (mercury) pressure having a duration of 75 msec. The CS overlapped the onset of the UCS, and they terminated at approximately the same time. The CS-UCS interval of 525 msec. and the duration of all stimuli were controlled by Hunter decade timers. The initiation of all trials was automatically programed with a Gerbrands programmer.

Procedure.—Each *S* was seated in a wooden armchair in his booth so he could not see any other *S*. After the headset had been attached, the instructions were read over the earphones. Conditioning trials began immediately and were spaced randomly at 15-, 20-, and 25-sec. intervals. The instructions were neutral and simply told *S* to relax and respond naturally. No instructions were given after the first session. Although *Ss* were not required to come at the same time each day, the daily sessions always were separated by a night's rest. Following the final session, each *S* filled out a questionnaire in which he was asked to give his impressions of the experiment.

Conditioned responses were recorded as deflections in the record greater than 1 mm. in the interval 150–500 msec. following CS onset.

RESULTS

Acquisition.—The mean number of CRs for the three groups during the first 50 trials was 31.6, 23.8, and 26.7 for Groups I, V, and X, respectively. These means were not significantly different, F (2, 54) = 1.10, owing to the large MS_e (269). Since it was not possible to assign *Ss* to conditions at random, there was also the possibility that the biased assignment would have serious effects on the analysis of asymptotic data. Consequently, the results from the first 50 trials were used to select two samples of *Ss* from each group. One sample consisted of "good" conditioners whose scores ranged from 30 to 50 CRs; the other sample consisted of "poor" conditioners whose scores ranged from 1 to 30. In order to match approximately good and poor conditioners across groups, and to keep sample size identical, it was necessary to drop three *Ss* from each leaving eight *Ss* in each sample in each group. The dropped *Ss* were not eliminated at random, but were selected in order to keep the mean number of CRs approximately the same for all three groups at a given level. The analysis by initial performance levels was done for two reasons: First, it provided rather gross matching of the groups and, hence, reduced some of the inter-*S* variability. Secondly, there was the possibility that the extremely high levels of performance obtained by some *Ss* might override possible significant findings.

Asymptotic performance.—Figure 1 presents the percentage of CRs by 25 trial blocks from Trials 101 to 500. Good and poor conditioners are shown separately for each experimental condition. The picture is a rather com-

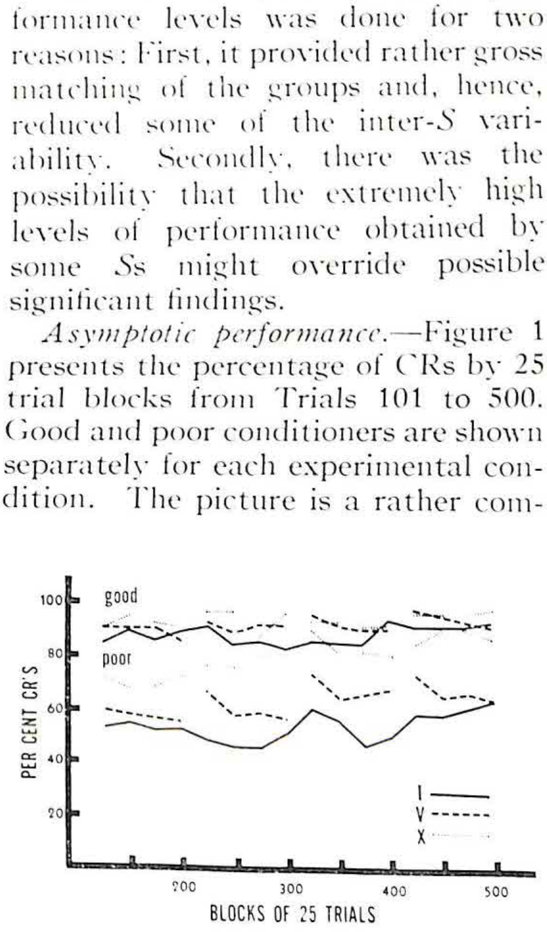


FIG. 1. Percentage of CRs in 25 trial blocks from Trials 101 to 500 for good and poor conditioners in each group.

plex one and can be understood best by referring to the statistical evaluation. Although it was possible to analyze the results of this experiment as a factorial design, the differences in number of trials per session did not make such a procedure advisable. A session for Groups I and V may be considered to be 100 trials, separated by a 10-min. and a 24-hr. rest, respectively, but a session for Group X was 50 trials. Thus, the results for each group were analyzed separately in the form of a Sessions×Performance Levels × Trials factorial analysis. Since this procedure increases the likelihood of accepting chance differences, some conservatism is advisable in interpreting the findings, particularly with respect to comparisons between groups. The results of these analyses are summarized in Table 1.

The main effects of performance levels and sessions were significant in all three groups. Although the levels effect is of no theoretical importance, the significant sessions effect indicates reliable performance increases over sessions for all groups. These changes interacted with performance levels in Groups V and X but not in Group I. The interaction took the form of greater increase in response frequency over sessions for the poor *Ss* as compared with the good *Ss*, particularly in Group X, where the curves for good and poor *Ss* converged by the last session as these *Ss* approached maximum possible performance. The only significant change within sessions occurred in Group V and was manifested as the main effect of trials. Inspection of the data revealed that this effect consisted of a decrease in frequency of responding within the 100-trial sessions.

Further explication of the within-sessions effect in Group V may be made by analyzing the trend over trials within each of the performance-levels groups for each session separately. Although the lack of significant interaction of sessions with trials makes it difficult to statistically justify such a procedure, the results are nevertheless enlightening and, for the most part, consistent with the picture shown in Fig. 1. For the good conditioners a significant negative linear slope was obtained only for Trials 401–500, F (1, 21) = 7.08. For the poor conditioners, significant negative slopes were obtained for Trials

TABLE 1

ANALYSIS OF VARIANCE OF WITHIN-SESSION EFFECTS FOR EACH GROUP

Source	Group I		Group V		Group X	
	df	*F*	*df*	*F*	*df*	*F*
Levels	1	8.81*	1	7.23*	1	5.46*
Error (B)	14	(541)	14	(464)	14	(133)
Sessions	3	6.80**	3	11.98**	7	5.21**
Trials	3	<1.00	3	6.87**	1	<1.00
S × L	3	<1.00	3	3.70*	7	3.46**
S × T	9	<1.00	9	<1.00	7	<1.00
L × T	3	<1.00	3	<1.00	1	1.16
S × L × T	9	1.07	9	<1.00	7	<1.00
Error (W)	210	(9.1)	210	(3.9)	210	(6.7)

* $p < .05$.
** $p < .01$.

201–300, F (1, 21) = 5.87, and for Trials 401–500, F (1, 21) = 8.06. The remaining sessions yielded Fs near or less than 1. The only exception to a clear interpretation is the failure of a consistent decrement to appear over Trials 301–400 for the poor conditioners. Although there was considerable decrement in this session in the first 50 trials, there was an inexplicable rise in performance late in the session which reduced the average slope. This improvement probably can be attributed to random error, however, since the F (2, 21) for deviations from linearity was only 2.36, $p > .10$. None of the other deviations even approached this level.

Not all *S*s showed evidence of decrements within sessions as defined by a negative slope coefficient for the blocks within a trial. In Group I, the number of *S*s showing a within-session loss was 6, 6, 8, and 9 for Sessions 2–5, respectively. For Group V, these values were 11, 9, 9, and 13. For Group X there were eight sessions, with the largest number of *S*s showing a loss being nine in Session 4. Generally, there were more such *S*s among the poor conditioners, but differences were small. Of even more significance is the fact that most of the *S*s in Group V who showed a loss within a session also showed a loss during other sessions. Of the 11 *S*s who showed a loss in Session 2, 7 also showed a loss in Session 3. Of the nine *S*s who showed decrements during Session 3, eight also showed losses during the succeeding session, while eight of the nine *S*s whose performance decreased during Session 4, also showed this pattern during Session 5. Furthermore, of seven *S*s who showed decrements during both Sessions 2 and 3, six also showed losses during Session 4. Of the eight *S*s who showed decrements in both Sessions 3 and 4, seven also showed decreases in performance during Session 5. None of the sessions in either of the other groups showed this consistency. In no case did the percentage of *S*s showing decrements in two successive sessions exceed 50%. In summary, then, some consistency was evidenced in that most of the *S*s in Group V who showed decrements in earlier sessions generally continued to do so. This was not true in the other groups where decremental performance during a session was unpredictable on the basis of previous performance.

The differences in overall performance among the three groups were quite large, particularly in the poor *S*s; however, statistical evidence for any difference was quite weak. Analysis of variance of the data in terms of blocks of 100 Trials×Groups ×Levels resulted in an F (2, 42) of 2.80 ($MS_e = 1516$) for the main effect of groups, which was just short of the conventional .05 confidence level, but obviously was reduced by near ceiling performance of the good conditioners in all groups. The interaction F for Groups×Levels, however, was near unity. Despite this lack of statistical support, the overall pattern of results in Fig. 1 tends to support the reliability of group differences.

In short, the pattern of results appears to be a combination of permanent depression in Group I, increases in frequency of responding over the 24-hr. rest interval in Group V with a decrease nearly to the level of responding of Group I within sessions, and consistent increases from session to session in Group X with no evidence of within-session decrements. The picture is clearer in the poor conditioners, but this may be due to a ceiling effect. In general, the pattern is similar at both performance levels.

Although the major analysis was performed on the two levels, an overall analysis based on all 19 *Ss* in each group resulted in essentially the same picture. Because the good conditioners performed at such a high level and tended to reduce any effects, the picture is not nearly as clear. Nevertheless, the combined results also showed significant decrements only in Group V.

Latencies.—Inspection of the distribution of CR latencies revealed that the increase in responding which was responsible for reminiscence in Group V consisted of a marked increase in short latency (150–250 msec.) responses. Within a session, however, these short latency responses disappeared, thus allowing frequency of response to return to approximately the prerest level. An overall picture of the changes in CR latency can be seen in Fig. 2, where the mean latency in blocks of 25 trials is plotted for Trials 101–500 for both good and poor conditioners in each group. Two things are evident from this graph: (*a*) although good conditioners tended to show shorter latencies than poor conditioners, the pattern within sessions across groups was similar; and (*b*) in general, mean latency increased within a session for all groups, but the increases were more consistent in Group V. Group V also showed consistently large reduction in mean latency over the 24-hr. rest. Others groups also showed this pattern but not as consistently, nor were the changes as great when they did occur.

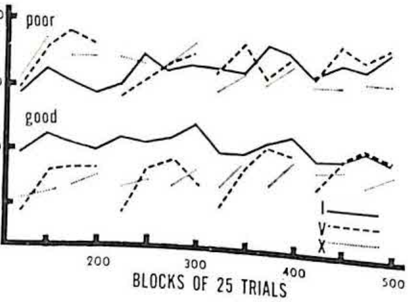


FIG. 2. Mean latency in 25 trial blocks from Trials 101–500 for good and poor conditioners in each group.

No statistical evaluation of these trends is presented since the means for various *Ss* are based on differing numbers of CRs and since there are occasional "empty" cells when an *S* made no CRs in that block of trials. The latency results are presented merely to illuminate the processes responsible for the results on frequency of response presented earlier in this section. The overall pattern of these findings is consistent with the frequency results, however, which were statistically significant.

DISCUSSION

The major finding of this study was that clear postasymptotic decrements in CR frequency appeared only in Group V and, even then, were small in magnitude. Concomitant lengthening of the latency of CRs within a session, followed by recovery with rest, supported this finding and, indeed, may represent the major inhibitory effect. This pattern also might be due to the diminution and recovery of a particular class of CRs, those with short latency.

There is no ready explication of the processes responsible for the decrements at the present time. It would appear, however, that conditions leading to high levels of responding and to short latencies would reduce the likelihood of finding such decrements. Certainly, the similarity between these results and performance curves obtained in studies of repetitive motor skills suggests an interpretation in terms of reactive inhibition or some similar concept. However, other processes, e.g., adaptation to the UCS or vigilance effects, would produce the same results.

REFERENCES

Ebel, H. C., & Prokasy, W. F. Classical eyelid conditioning as a function of sustained and shifted interstimulus intervals. *J. exp. Psychol.*, 1963, **65**, 52–58.

Krauss, H. H., & Prokasy, W. F. On intertrial interval discrimination in classical conditioning. *Psychol. Rep.*, 1963, **12**, 158.

Pavlov, I. P. *Conditioned reflexes.* London: Oxford Univer. Press, 1927.

Prokasy, W. F., Ebel, H. C., & Thompson, D. D. Response shaping at long interstimulus intervals in classical eyelid conditioning. *J. exp. Psychol.*, 1963, **66**, 138–141.

Prokasy, W. F., & Whaley, F. L. The intertrial interval variable in classical conditioning. *J. exp. Psychol.*, 1961, **62**, 560–564.

Prokasy, W. F., & Whaley, F. L. Intertrial range shift in classical eyelid conditioning. *Psychol. Rep.*, 1963, **12**, 55–58.

(Received July 9, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 5, 526–533

INTERNAL CONSISTENCY OF SUBJECTIVE PROBABILITIES[1]

CAMERON R. PETERSON,[2] Z. J. ULEHLA,[3]
ALAN J. MILLER, LYLE E. BOURNE, JR.
University of Colorado
AND DONALD W. STILSON
University of Colorado School of Medicine

3 experiments evaluated the degree to which relations within sets of subjective probabilities conformed to implications of mathematical probability theory. According to probability theory, 2 products of subjective probabilities, $\psi(a)\psi(b|a)$ and $\psi(b)\psi(a|b)$, should be equal. The mean correlation between these products was .67 in Experiment I. The corresponding correlation was .90 in Experiment II after partialing out the effect of relevant objective probabilities. In Experiment III, Ss made trial-to-trial revisions of subjective probabilities which referred to 2 classes of events. Although the magnitude of subjective-probability revision was less than the magnitude of objective-probability revision, the relation between revised subjective probabilities associated with the 2 classes of events agreed closely with the relation specified by probability theory.

Two important characteristics of subjective probabilities can be distinguished—veridicality and consistency. The subjective probability of X, $\psi(X)$, is veridical to the extent that its value agrees with the corresponding objective probability, $P(X)$. A set of subjective probabilities is internally consistent to the extent that the set conforms to implications of mathematical-probability theory. That is, if an equation is derived from probability theory and subjective probabilities are substituted in the equation, then the subjective probabilities are internally consistent to the extent that the two sides of the equation agree. Since objective probabilities are consistent with probability theory, perfect veridicality implies internal consistency of subjective probabilities; the reverse is not true, however.

The experiments reported herein evaluated the internal consistency of sets of subjective probabilities. As referents for subjective probabilities, Exp. I used naturalistic events with unknown objective probabilities and Exp. II used events with *E*-determined probabilities. Objective probabilities of the referents in Exp. I and II remained fixed throughout the experiment and will henceforth be termed "static." Experiment III investigated the revision of subjective probabilities and used as referents two classes of E-determined events, data and data generators.

---

[1] Experiment I was conducted by Donald W. Stilson at the University of Colorado Medical School and was supported by Research Grant M-5744 from the National Institute of Mental Health. Our thanks to A. Omer Dery for collecting the data and helping with the analysis. Experiment II was conducted by C. R. Peterson and Z. J. Ulehla, and Exp. III was conducted by C. R. Peterson, A. J. Miller, and L. E. Bourne, Jr. Both were performed at the Behavior Research Laboratory, Institute of Behavioral Science, University of Colorado, and are combined as Publication No. 52 of the Institute. The research was supported by Research Grant M-4977 from the National Institute of Mental Health. The authors are indebted to K. R. Hammond and F. J. Todd for helpful suggestions.

[2] Now at the University of Michigan.

[3] Now at the University of Denver.

EXPERIMENT I

If $P(a)$ represents the probability of event a and $P(a|b)$ is the conditional probability of a given b, then it follows from the definition of conditional probability that

$$P(b)P(a|b) = P(a)P(b|a). \qquad [1]$$

Each side of Equation 1 equals the probability of the joint occurrence of a and b.

Experiment I measured subjective probabilities corresponding to each of the four probabilities in Equation 1. Then, over a set of observations for each S, a linear correlation was calculated between one product of subjective probabilities, $\psi(a)\psi(b|a)$, and the other product, $\psi(b)\psi(a|b)$. Degree of consistency was inferred from degree of linear dependence between the two products.

Method

Subjects.—Twelve introductory psychology students served as *S*s in order to satisfy a course requirement.

Apparatus and procedure.—Response booklets contained subjective-probability scales. The referents for subjective probabilities were 20 different trait names such as GOOD, DISHONEST, WITTY, and WEAK. The *S*s first responded with unconditional probabilities for each of the trait names, (e.g., "How many people in 100 could be described correctly using the word . . ."), and then with conditional probabilities for each trait name given each of the other trait names. Conditional-Probability scales were of the following form:

> "One hundred persons are known to be weak. How many of them would you expect to be . . .
>
> 0 10 20 . . . 100
>
> (a) DULL . . . .
>
> (b) BRAVE . . . . ," etc.

TABLE 1

CONSISTENCY AND RELIABILITY CORRELATIONS FOR EACH *S* IN EXP. I

S	Consistency	Reliability
1	.69	.83
2	.49	.79
3	.48	.65
4	.31	.68
5	.56	.66
6	.91	.91
7	.78	.57
8	.72	.68
9	.56	.77
10	.68	.72
11	.89	.68
12	.45	.53
Mean	.67	.72
Upper 99% Limit	.80	.80
Lower 99% Limit	.49	.62

Each pair of traits furnished two subjective-probability products, of the form of Equation 1 [e.g., ψ(WITTY)ψ(BRAVE|WITTY) and ψ(BRAVE)ψ(WITTY|BRAVE)], for a total of 190 bivariate observations. A product-moment correlation computed from these observations served as the measure of each *S*'s internal consistency. The *S* gave two different estimates, separated in time, of the unconditional probabilities thus permitting an assessment of reliability.

Results

The hypothesis of consistency explicitly requires an intercept of 0 and a slope of 1.0 for each *S*, when the aforementioned subjective-probability products are plotted. Inspection of the individual scatter diagrams revealed no obvious departure from these values, thus leading to the acceptance of the correlation coefficient[4] as an adequate statistic for evaluating consistency.

As shown in Table 1, the mean consistency correlation is .67 with 99% confidence limits of .49 and .80. These results indicate a moderate degree of internal consistency. The mean reliability correlation between the first and second estimates of unconditional probabilities is only .72, a value which places a rather low upper limit upon consistency.

---

[4] Statistical computations upon correlation coefficients have used the Fisher *z* transformation.

EXPERIMENT II

Experiment II investigated the same basic problem as Exp. I, but with different apparatus, different *S*s, and different referents for the subjective probabilities. The referents were well-defined events with which each *S* had considerable experience, and the objective probabilities of these events were experimentally controlled. Control of objective probability enabled the contribution of veridicality to the subjective probabilities to be partialed out in the evaluation of internal consistency. It is possible that $\psi(a)\psi(b|a)$ may approximate $\psi(b)\psi(a|b)$ simply because the subjective probabilities are accurate. The subjective products would be correlated with each other if they were correlated with the corresponding objective values. Since objective probabilities were known, the correlation of $\psi(a)\psi(b|a)$ with $\psi(b)\psi(a|b)$ was computed with the dependence of these products on the corresponding objective probabilities partialed out. This positive partial correlation was taken as a reflection of the extent of internal consistency that was not a by-product of the veridicality of the subjective probabilities.

Method

Subjects.—Sixteen paid volunteers obtained through the University of Colorado Student Employment Service served as *S*s.

Apparatus.—The response apparatus for obtaining subjective-probability estimates was a 25-in. bar divided into 100 equal units which were numbered from left to right. Three sliding markers permitted *S*s to partition the bar into four sections. The relative lengths of the respective sections represented *S*'s subjective probabilities of the associated events. The length of the entire rod represented a probability of 1.0, so subjective probabilities for the four events were required to sum to 1.0.

Procedure.—During an earlier experiment, *S*s were trained in a conditional-probability learning task. The task included 128 possible stimulus-response (a_ib_j) pairings. During each trial *S* observed which a_i was on the front of a card, predicted which b_j would be on the back, and then observed which b_j actually was on the back. After each block of trials *S*s used the response bar for selected unconditional- and conditional-probability estimates of the a_i and b_j events (for more detail see Peterson & Ulehla, 1964). Participation in the earlier experiment thus served to provide *S*s with (*a*) a clear definition of the a_i and b_j events, (*b*) some knowledge of the relative frequencies of the events, and (*c*) experience in estimating both unconditional and conditional probabilities. This participation prepared *S*s for Exp. II in which they were required to estimate all possible unconditional and conditional probabilities of the events a_i and b_j. Insertion of these subjective probabilities for each *S* into Equation 1 yielded 128 bivariate observations of the appropriate form, $\psi(a_i)\psi(b_j|a_i)$ and $\psi(b_j)\psi(a_i|b_j)$.

Results

As Table 2 shows, the product-pair consistency correlations are high. The mean is .92 with 99% confidence limits of .88 and .94. On the other hand, the product pairs show low correlations with the objective probabilities; the mean being .34. Thus, the consistency partial correlations with dependence upon objective values partialed out are also high, with a mean of .90 and 99% confidence limits of .86 and .93. Across *S*s, consistency is positively correlated with veridicality; the correlation between "consistency" and "veridicality" coefficients is .69 ($p < .01$). In sum, *S*s exhibited a high degree of internal consistency and the more consistent *S*s also tended to be the more veridical.

TABLE 2

CONSISTENCY AND VERIDICALITY CORRELATIONS FOR EACH *S* IN EXP. II

S	Consistency	Veridicality	Consistency Veridicality
1	.93	.47	.91
2	.91	.48	.89
3	.96	.41	.95
4	.92	.12	.92
5	.89	.06	.89
6	.88	.05	.88
7	.91	.45	.90
8	.89	.36	.88
9	.95	.62	.92
10	.91	.21	.91
11	.96	.20	.96
12	.96	.64	.93
13	.96	.60	.95
14	.89	.28	.88
15	.67	−.03	.67
16	.80	.16	.80
Mean	.92	.34	.90
Upper 99% Limit	.94	.47	.93
Lower 99% Limit	.88	.18	.86

EXPERIMENT III

Experiments I and II demonstrated that *S*s show a degree of internal consistency in their assignment of static subjective probabilities to interrelated events. Experiment III was designed to discover if *S*s would show an analogous consistency when they revised their subjective probabilities upon the occurrence of relevant data.

Bayes' theorem applied to objective probabilities specifies the theoretically correct revision of these probabilities as the result of information provided by the occurrence of a datum (Edwards, Lindman, & Savage, 1963). Several experiments have used Bayes' theorem as a normative model for evaluating the veridicality of subjective-probability revision. In those experiments, *S*'s task was to use the information provided by the occurrence of a datum (D) to revise his subjective probability about which hypothesized data generator (H) was responsible for generating the D. The typical finding has been that the $\psi(H)$s were revised less than the corresponding $P(H)$s (Edwards, Lindman, & Phillips, 1965; Edwards & Phillips, 1964; Peterson & Miller, 1965; Peterson, Schneider, & Miller, 1965). Subjective revision has usually been more conservative than veridical revision.

Even though subjective-probability revisions are conservative, the results of Exp. I and II suggest that a set of revised subjective probabilities may be internally consistent. Deviations from veridicality for one subjective probability may elicit compensatory deviations for related subjective probabilities. Because the evaluation of internal consistency requires the revision of subjective probabilities associated with at least two events, Exp. III extended the paradigm used in the aforementioned studies on subjective-probability revision by investigating the revision of $\psi(D)$ as well as of $\psi(H)$.

Where H_a and H_b are two mutually exclusive and exhaustive Hs, any $P(D)$ can be partitioned as follows (Edwards et al., 1963, p. 200):

$$P(D) = P(H_a)P(D|H_a) + P(H_b)P(D|H_b). \quad [2]$$

Since H_a and H_b are mutually exclusive and exhaustive, $P(H_b)$ is equal to one minus $P(H_a)$. By making this substitution into Equation 2 and simplifying,

$$P(D) = P(D|H_b) + [P(D|H_a) - P(D|H_b)]P(H_a). \quad [3]$$

Equation 3 provides a linear expression relating $P(D)$ to $P(H_a)$.

To evaluate consistency, obtained subjective revisions of $\psi(H_a)$ and $\psi(D)$ were substituted for $P(H_a)$ and $P(D)$ in Equation 3 (remaining terms were controlled experimentally). To the extent that the two sides of the equation agree, revisions of $\psi(H_a)$ and $\psi(D)$ are internally consistent. To evaluate veridicality, objective revisions in $P(H_a)$ over a sequence of trials were calculated by means of Bayes' theorem. Substitution of these objective revisions into Equation 3 yielded revisions of $P(D)$. To the extent that subjective revisions in $\psi(D)$ agree with objective revisions in $P(D)$, subjective-probability revision is veridical.

Method

Experimental design.—The experiment was patterned after an earlier experiment on subjective-probability revisions (Peterson & Miller, 1965). The design involved two data generators, H_a and H_b, and two types of data, D_b and D_w. On the basis of information provided by the display of one of the Ds, S revised his subjective probability about whether it was generated by H_a or by H_b. In addition, S also revised subjective probabilities about which D would occur on the next trial. Thus, after observing the D, S first revised $\psi(H_a)$ and $\psi(H_b)$, and then revised $\psi(D_b)$ and $\psi(D_w)$.

The experiment consisted of 15 blocks of 10 trials each. Each block of trials represented a prearranged sequence of 10 data. The order in which sequences were assigned to blocks was counterbalanced across Ss within Blocks 1–5, Blocks 6–10, and Blocks 11–15. In order to investigate the effect upon internal consistency of making joint subjective-probability estimates of both Hs and Ds, the first set of 5 blocks was a control set. During that set, Group I Ss responded with $\psi(H)$s and $\psi(D)$s; Group II responded with only $\psi(D)$s; and Group III responded with only the $\psi(H)$s. During all later blocks all groups responded with both $\psi(H)$s and $\psi(D)$s.

Stimulus sequences were arranged in advance in order to permit comparison of responses across Ss. Fifteen sequences of 10 data each were generated randomly within the constraint that they conform to a binomial distribution with probability of the most likely datum equal to .67. The "correct" H was selected randomly as H_a for seven of the sequences and as H_b for eight of the sequences. A positive (the most likely) outcome was D_b for H_a and D_w for H_b.

Subjects.—The 30 Ss were students from an introductory psychology course. Participation in the experiment satisfied a course requirement.

Apparatus.—Two six-sided dice served as Hs. The H_a had four black and two white sides. The H_b had two black and four white sides. The D_b represents a black side and D_w a white side. The S's subjective-probability estimates of D and H were made, on each trial, on different response bars similar to those used in connection with Exp. II. The response bar for the $\psi(H)$s ranged in value from .00 to 1.00; the response bar for the $\psi(D)$s ranged from .33 to .67, which were the limits dictated by the composition of Hs.

Procedure.—For each block of trials, E selected either H_a or H_b seemingly (to S) at random, and then progressed through the 10 trials. Prior to Trial 1 in each sequence, Ss were required to set probability markers at .50 on both response bars. The E rolled the chosen die and reported which side showed up. The reported side was actually determined by the presequencing rather than by the roll. The Ss revised and recorded $\psi(H_a)$ and $\psi(D_b)$. The E then rolled the die again, reported the side, and the Ss revised subjective probabilities. This procedure continued throughout each block of 10 trials.

With respect to the revision of subjective probabilities, Ss were instructed that a black side was more likely to be rolled by the "mostly black" die and a white side by the "mostly white" die. Therefore, a roll which came up black should increase their $\psi(H_a)$s and also their $\psi(D_b)$s. The amount of increase, however, depended upon each S's intuitive feeling of where the marker should go.

Results

Internal consistency.—An internally consistent $P(D)$ is related to a $P(H)$ by Equation 3. Substitution of the relevant experimental values of $P(D_b|H_a)$ and $P(D_b|H_b)$ into Equation 3 yields the linear equation,

$$P(D_b) = \tfrac{1}{3} + \tfrac{1}{3}P(H_a), \qquad [4]$$

with an intercept of $\frac{1}{3}$ and a slope of $\frac{1}{3}$.

In order to determine the linear relation between $\psi(H_a)$ and $\psi(D_b)$ product-moment correlations were computed over the last 100 trials for each S. The mean within-S correla-

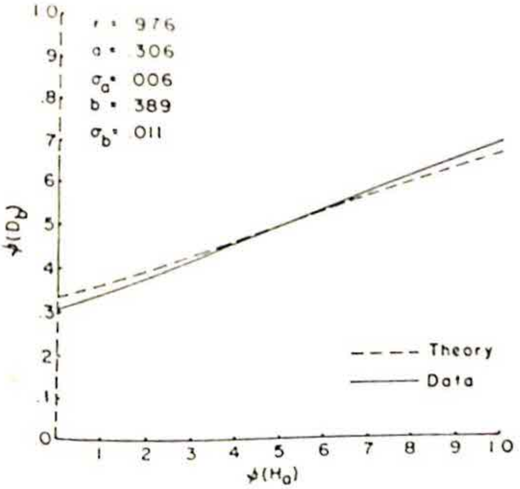


FIG. 1. A comparison of the theoretical and obtained $\psi(D_b)$ as a function of $\psi(H_a)$. (The $\psi(D_b)$ was restricted to the solid portion of the ordinate between .33 and .67. Values in the upper left hand corner refer to means and to standard error of means.)

tion across the 30 *Ss* was .976, indicating that a straight-line function accounts for about 95% of the covariation between $\psi(H_a)$ and $\psi(D_b)$. Accordingly, a least-square analysis was used to find the straight line of best fit for each *S*. The solid line in Fig. 1 is defined by the mean intercept (a = .306) and mean slope (b = .389) across the 30 *Ss*. The standard error of the mean for the intercept is .006 and for the slope is .011.

These results indicate that *Ss* deviated systematically from perfect consistency. The mean subjective intercept is more than four standard scores below the ideal and the mean subjective slope is about five standard scores above the ideal. Thus the rate of $\psi(D_b)$ revision with respect to $\psi(H_a)$ revision was significantly greater than that specified by Equation 4.

Even though the subjective data line is significantly different from the theoretical ideal, the actual magnitude of the difference is small. With the indifference point of (.5, .5) as an origin, the data line is rotated less than 3° counterclockwise from the theoretical line. Part of the reason for the closeness of these two lines may be that *Ss* were instructed to revise $\psi(H)$ and $\psi(D)$ in the same direction upon the observation of a datum. The amount of revision, however, was left up to each *S*'s own intuitive feeling. The instructions did not specify that the amount of $\psi(D)$ revision with respect to $\psi(H)$ revision should be constant across trials so the subjective values need not have provided such an orderly relation to any function. Also the instructions did not specify that the amount of $\psi(D)$ revision with respect to $\psi(H)$ revision be constant over all values of $\psi(H)$ so the subjective function need not have been a straight line so near the theoretical ideal. Obviously *Ss* provided considerable constraints of their own in the direction of internal consistency.

Veridicality.—Figure 2 presents the least-square line relating revisions in $\psi(D_b)$ to objective revisions in $P(D_b)$

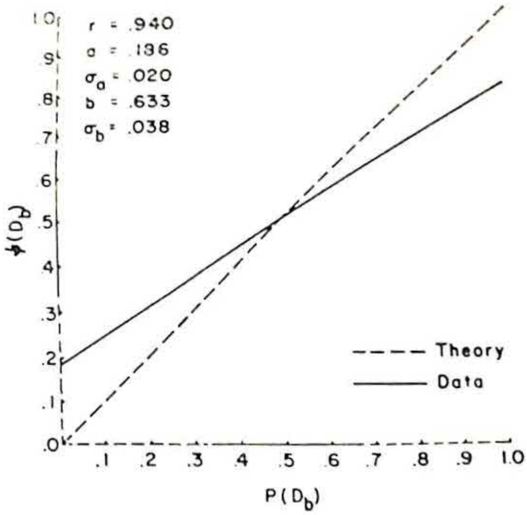


FIG. 2. A comparison of the theoretical and obtained $\psi(D_b)$ as a function of $P(D_b)$. (All probabilities were restricted to the solid portion of the coordinates between .33 and .67. Values in the upper left hand corner refer to means and to standard error of means.)

for the 30 *S*s during the last 10 blocks of trials. The mean within-*S* product-moment correlation of .94 suggests again that the straight-line analysis is appropriate for the data. The clockwise rotation of the empirical line from the theoretical line means that revision in $\psi(D_b)$ was more conservative than the correct revision. Subjective revision was positively correlated with, but of less magnitude than, the correct revision. Between-*S* variation, as indicated by the standard errors of the mean for the intercept and slope, is about three times larger for veridicality than for consistency.

Effect of joint responding. Two kinds of evidence relate to whether the high degree of internal consistency is a result of responding jointly with both $\psi(H)$ and $\psi(D)$. Table 3 presents comparative group measures of deviations from internal consistency. The group mean $\psi(H_a)$ for each of the 50 trials of the first five blocks was substituted into Equation 4 in order to predict the corresponding mean $\psi(D_b)$ in either the same or a different group. The mean difference and the σ_d refer to differences between predicted and obtained $\psi(D_b)$s. Zero difference implies perfect consistency. Since prediction error was not reliably less within Group I (Row 1) than for the other three comparisons (Rows 2, 3, and 4), the process of joint responding was not an important determiner of internal consistency.

TABLE 3

Mean Difference between Predicted and Obtained $\psi(D_b)$

$\psi(H_a)$	$\psi(D_b)$	Mean Difference	σ_d
Group I	Group I	.009	.001
Group I	Group II	.009	.001
Group III	Group I	.012	.001
Group III	Group II	.011	.001

Internal consistency was high whether or not *S*s made joint responses. Additional evidence for this point resides in the results of an analysis of variance within the last 10 blocks. Groups were not significantly different on either the consistency slope, $F(2, 18) = 0.19$, $p > .25$, or the consistency correlation, $F(2, 18) = 2.31$, $p > .10$. Thus, joint responding provided neither an immediate nor a long-term contribution to internal consistency.

Discussion

The major finding of these experiments is that a set of subjective probabilities is constrained toward internal consistency. This internal consistency characterizes both static subjective probabilities and the revision process. Revision consistency does not depend upon the process of joint responding.

Considerably stronger consistency was found in Exp. II and III than in Exp. I. Several experimental differences offer possible explanations for this discrepancy. The referents of the subjective probabilities in Exp. I were not as clearly defined as in Exp. II and III. It is possible that Trait a_1 may have a different meaning when associated with Trait a_2 than when associated with Trait a_3. There is no reason to expect internal consistency if the referents of the subjective probabilities do not remain constant. In addition, Exp. II and III imposed more stringent response constraints; the subjective probability response apparatus required each set of subjective probabilities to sum to 1. In Exp. I each subjective probability could take on any value between 0 and 1.0. Furthermore, instructions in Exp. III provided some lower bound for consistency. These constraints may have been partly responsible for the differential consistency.

The internal consistency of Exp. II was not due to veridicality; Exp. III found consistency of revision to be much stronger than veridicality of revision.

The direction of the departure from veridicality in Exp. III was in agreement with the earlier finding that subjective revision is conservative in relation to the correct revision. When *S*s are conservative on one revision, the degree of conservatism on the other revision is such that both revised subjective probabilities remain internally consistent. Although the diagnostic impact of the datum is too weak, it is consistently weak from one variable to another.

The significance of these results is that a set of subjective probabilities appears to form an integrated system. Departure from veridicality in one part of the system implies compensatory departure in other parts. Thus, individual probability estimates referring to related events probably reflect an underlying network of probabilistic relationships, rather than independently stored data. Since both estimates refer to the same probability system, the two resulting measures are interdependent.

REFERENCES

Edwards, W., Lindman, H., & Phillips, L. D. Emerging technologies for making decisions. In *New directions in psychology.* New York: Holt, Rinehart, and Winston, 1965. Pp. 259–325

Edwards, W., Lindman, H., & Savage, L. J. Bayesian statistical inference for psychological research. *Psychol. Rev.*, 1963, **70**, 193–242.

Edwards, W., & Phillips, L. D. Man as transducer for probabilities in Bayesian command and control systems. In M. W. Shelly & G. L. Bryan (Eds.), *Human judgments and optimality.* New York: Wiley, 1964. Pp. 360–401.

Peterson, C. R., & Miller, A. J. Sensitivity of subjective probability revision. *J. exp. Psychol.*, 1965, **70**, 117–121.

Peterson, C. R., Schneider, R. J., & Miller, A. J. Sample size and the revision of subjective probabilities. *J. exp. Psychol.*, 1965, **69**, 522–527.

Peterson, C., & Ulehla, Z. J. Uncertainty, inference difficulty, and probability learning. *J. exp. Psychol.*, 1964, **67**, 523–530.

(Received July 18, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 5, 534–536

SUPPLEMENTARY REPORTS

USE OF NEGATIVE SLOPE TRANSFORMATIONS OF KNOWLEDGE OF RESULTS ON A SIMPLE MOTOR RESPONSE[1]

DONALD A. SCHUMSKY[2]

Tulane University

Bilodeau (1953) studied a series of transformations of knowledge of results (KR) which had slope and intercept constants varying inversely with one another making it impossible to partition their separate effects. The present study, by employing transformations with both positive and negative slopes, had comparable groups equated for slope constant but varying in intercept value. 80 Ss were given 16 discrete trials to learn a simple knob-turning response. Each response was followed by an appropriately transformed KR. Results revealed that S's initial responding is related to the value of the intercept constant of the transformation equation. Subsequently, the speed of learning is directly related to the absolute size of the slope constant of the transformation equation.

In 1953, Bilodeau studied the effects on performance of a number of linear transformations of the form $Y = aX + b$. He varied the slope and intercept constants on these transformations while holding goal score equal (all transformation lines intersected at the goal). He reported his results as suggesting "A family of curves" . . . of performance . . . "each member with the same origin and apparently the same performance asymptote. Using accuracy of response as a criterion of learning, the groups are ranked with the one with greatest slope . . . and smallest intercept . . . to the one with the least slope and largest intercept [p. 413]." The problem that confronted Bilodeau in the interpretation of his data was that in his study the values of the slope and intercept constants varied inversely with one another. It was thus impossible to attribute particular learning effects to the variation of either slope or intercept constants.

The main purpose of the present study, then, was to investigate further the relationships between *S*'s response and various parameters of the transformation equations. By employing transformations with different slope direction (positive or negative sign) it was possible to have two groups with the same growth constant (absolute value of slope constant) and the same goal score. Such a set of transformations would yield pairs of groups equated for the value of growth constant but which varied in intercept value. Also, if true score were plotted on the x axis and KR plotted on the y axis each pair of lines with the same growth constant would be symmetrical in terms of deviation from the goal along the y axis. Since all transformation lines intersect at a common goal score, the discrepancy between reported score and the goal score for any given response value (this affording the subjective estimate of error) was held constant for each pair of groups having the same absolute value of growth constant.

Method.—The apparatus, instructions, and procedures in this study were the same as those described in Bilodeau (1953) except for the following modifications:

1. There were 20 *S*s in each group. They were undergraduate psychology students, and both males and females were used.
2. All *S*s were instructed to make six half turns on the first trial.
3. Their target score was defined at 250.
4. There were four experimental groups in this study, identified by sign of slope (P = positive, N = negative) and value of the growth constant. The four groups and their transformation equations were:[3]

---

[1] This is based on a thesis submitted to the Psychology Department of Tulane University, in partial fulfillment of the requirements for the MS degree. The author wishes to express his indebtedness to E. A. Bilodeau, A. L. Irion, and C. W. Mann for their advice during the course of this study.

[2] Now at the University of Cincinnati.

[3] The slope constant values were chosen from among those used by Bilodeau (1953). Thus the two positive groups (P-12.5, P-75) are repeats of values he had investigated with different intercept values.

Group	Equation
P-12.5	$Y = +0.125\ X + 218.75$
N-12.5	$Y = -0.125\ X + 281.15$
P-75	$Y = +0.75\ X + 62.5$
N-75	$Y = -0.75\ X + 437.5$

Results.—All scores were converted into response-error scores by the formula $Y = X - 250$. The mean absolute error for each group can be seen plotted as a function of trials in Fig. 1. It can be seen from this figure that the functions obtained from the present study resemble members of Bilodeau's "family of curves." These curves all show a general trend downward, though as expected the two groups with negative slope transformations show an initial trend in an upward direction. The points of inflexion downward toward a reduction of error occur at Trial 2 for Group N-75, and Trial 3 for Group N-12.5. The groups with the positive slope transformations show marked superiority in speed of attainment of the goal response. At Trial 2 the rank order of all groups is directly related to the size of the intercept constant and the groups were significantly different from each other (Kruskal-Wallis $H = 57.80$, $p < .001$). The two negative groups cross between Trial 2 and Trial 3 and throughout the remainder of practice both groups with slope 75 demonstrate performance superior to their respective controls. By Trial 16 all groups except N-12.5 seem to have reached the same level of performance.

Bilodeau's (1953) suggestion that *S*s initially respond as if there is a positive relationship between true score and KR is supported by the behavior of the negative transformation groups. These groups showed an initial increase in response error after Trial 1 (the first KR). The point of inflexion downward (toward decreased error) can be considered the average point at which the erroneous hypothesis was rejected.

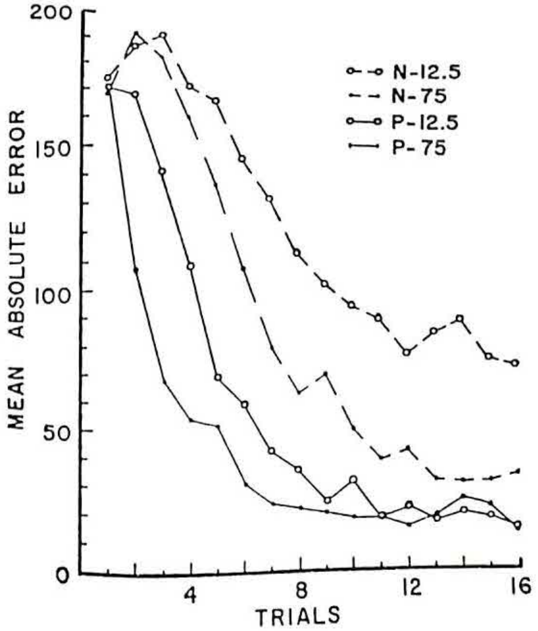


FIG. 1. Mean absolute error as a function of trials for all experimental groups.

The standard deviations of the true-score values were examined. In general, all groups showed an initial marked increase in variability followed by a general decrease. It may be noted that Bilodeau's finding, that the trial at which the point of maximum variability occurred is a function of experimental treatment, did not completely obtain here. The points of maximum variability for the two positive groups fell at Trial 3, for Group N-75 at Trial 5, and for Group N-12.5 at Trial 14.

Discussion.—Bilodeau (1953) has stated that the parameters and signs of the transformation equations define the relative difficulty of *S*'s problem. Given that *S* responds initially as if he expects the relationship between true score and reported score to be $Y = 1.0\ X + 0$, the degree to which a given transformation deviates from $Y = 1.0\ X + 0$ should define the difficulty of the problem. Since the value of the intercept constant varies directly as a function of the degree by which a line deviates from $Y = 1.0\ X + 0$ along the y axis (reported score), we should find that the difficulty presented by a transformation should vary initially as a function of the intercept constant, goal score held equal. This would seem to be supported by the data in Fig. 1 where mean absolute error on Trial 2 seems to be monotonically related to increases in the value of the intercept constant. In addition, the speed with which this hypothesis was rejected would vary as a function of the increment in reported score per unit response (slope constant). From Trial 3 on, one may suspect that some *S*s have already rejected the erroneous hypothesis, the speed of rejection of the hypothesis being a function of the slope constant (the two negative groups cross).

The *S*'s speed of acquisition of the goal response should be a function of his estimate of error. The *S*'s estimate of error must be based on the discrepancy between his KR and the goal. In the present study the groups with slope of 75 were the groups receiving KRs most removed from the goal score. A comparison of each of these groups with the 12.5 group having the same slope direction indicates that the speed of acquisition is indeed related to the reported score-goal score discrepancy.

Following the line of reasoning suggested by Bilodeau (1953), a great spread of scores would indicate that a number of different hypotheses were being tried by the individuals in the groups. He suggested that this number increased with the difficulty of the problem presented by the given transformation equation. In general this thinking was supported by the data of the present study.

REFERENCE

Bilodeau, E. A. Speed of acquiring a simple motor response as a function of the systematic transformation of knowledge of results. *Amer. J. Psychol.*, 1953, **66**, 409-420.

Received June 18, 1964

Journal of Experimental Psychology
1965, Vol. 70, No. 5, 536-537

SEX DIFFERENCES IN ADAPTATION OF THE GSR UNDER REPEATED APPLICATIONS OF A VISUAL STIMULUS[1]

H. D. KIMMEL AND ELLEN KIMMEL

University of Florida

8 men and 8 women came to the laboratory on 3 occasions, 1 wk. apart, to receive 10 presentations of a visual stimulus on each occasion. The average GSR to the light showed an intrasession (but no intersession) adaptation effect for both sexes but this effect was almost gone for the males on the last session. In addition, the males made significantly larger GSRs than the women on all 3 sessions, especially on the last session. It was conjectured that the men became familiar with the procedure and anticipated the end of the 3rd session by making larger GSRs, but that the women did not.

Kimmel (1964) recently reported data on the magnitude of the GSR to a visual stimulus presented repeatedly in a single laboratory session. The present report extends the earlier procedure to one in which *Ss* returned to the laboratory on two later occasions, each 1 wk. apart, to receive further presentations of the same stimulus. In addition, a possible sex difference in GSR responsiveness was considered.

Method.—Sixteen *Ss*, eight men and eight women, volunteered to come to the laboratory for three separate sessions, each 1 wk. after the previous one. On each session *Ss* sat in an air-conditioned, sound-proof, double-wall IAC chamber for 20 min. without stimulation and then received 10 1-sec. presentations (intertrial interval = 20–30 sec.) of a dim circular red light produced by a Grayson-Stadler multiple-stimulus projector. The palmar GSR was picked up by zinc-zinc sulphate-saline electrodes, amplified by a Biophysical Instruments Company (chopper type) amplifier, and recorded on a Texas Instruments Company Recti-Riter with a paper speed of 12 in/min. All *S* had to do was relax and pay attention.

Results.—Figure 1 shows the average GSR magnitude of the males and females on each of the 10 trials on each session. It is clear in the figure that adaptation (a drop in average GSR magnitude) occurred on all three sessions for both sexes (but not from session to session), and on the third session, the males first dropped and then increased toward the end of the session. On all three sessions, the males made larger responses to the light than did the females, especially on the third session. Analysis of variance showed that the overall sex difference was significant, $F\ (1, 14) = 4.87$, (error $MS = .59$), the Trials (adaptation within sessions) effect was significant, $F\ (9, 126) = 14.23$, (error $MS = .03$), and the interaction of Trials × Sessions was significant, $F\ (18, 252) = 1.65$, (error $MS = .04$). The last-mentioned effect, of course, reflected the fact that adaptation on the last session's trials was not nearly as noticeable as it had been on the two previous sessions. Neither the Trials, the Sessions, nor the Trials × Sessions effects interacted significantly with Sex, although the data suggest a tendency in this direction.

Discussion.—While these results generally confirm the earlier conclusions regarding adaptation of the GSR to an innocuous stimulus, two additional observations are in order. First, males respond at a higher level than females in this type of situation. This finding is not unlike earlier observations of a sex difference in GSR responsiveness (e.g., Kimmel & Hill, 1961). Secondly, males tended to show less adaptation in later sessions, particularly on the last session. What this may have been due to is not immediately

---

[1] Done under Grant NIH-MH-06060-02. Participation of the second author was supported by a Predoctoral Research Fellowship PHS-MH-16798-02.

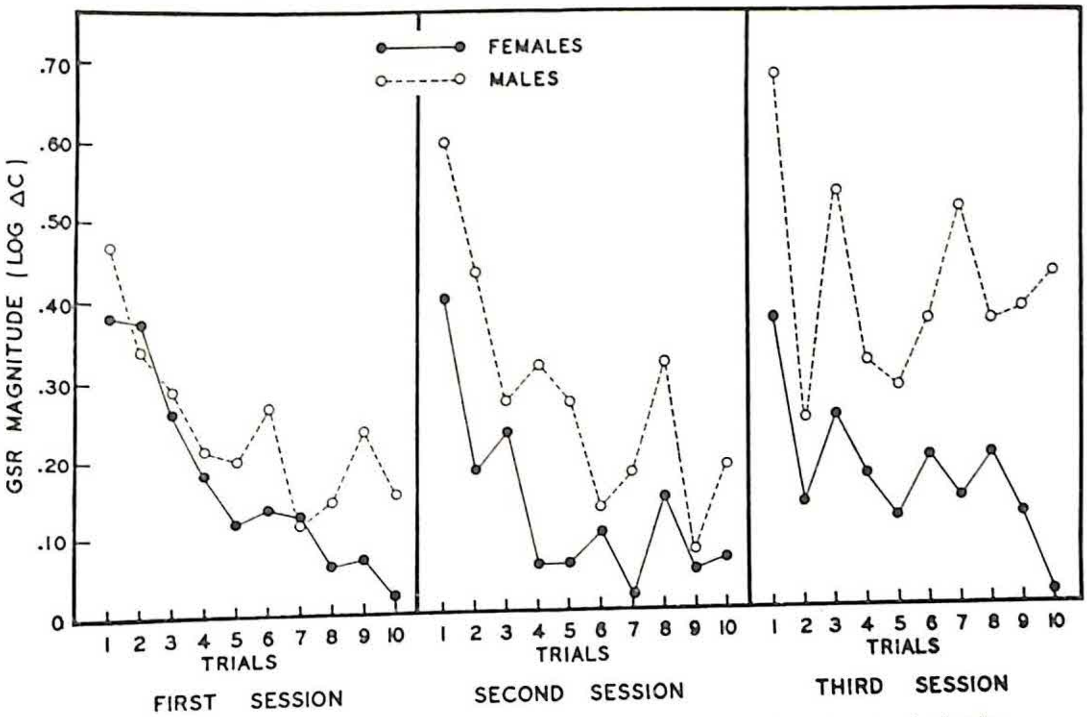


FIG. 1. Mean GSR magnitude (of men and women) on 10 presentations of a visual stimulus on each of three sessions.

apparent, but it may be conjectured that familiarity with the procedure, gained during previous sessions, may have inclined the males to anticipate the conclusion of the third session and, thus, to become more responsive. Even if this conjecture is feasible, however, it does not explain why only the males responded in this way, other than that females are less likely to adjust in this way. It would appear that further work on sex differences in GSR responsiveness is necessary.

REFERENCES

KIMMEL, H. D. Adaptation of the GSR under repeated applications of a visual stimulus. *J. exp. Psychol.*, 1964, **68**, 421–422.

KIMMEL, H. D., & HILL, F. A. Two electrodermal measures of response to stress. *J. comp. physiol. Psychol.*, 1961, **54**, 395–397.

(Received June 29, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 5, 537–539

COMPARISON OF PAIRED-ASSOCIATE TRANSFER EFFECTS BETWEEN THE A-B, C-A AND A-B, B-C PARADIGMS[1]

L. R. GOULET[2] AND A. BARCLAY

Saint Louis University

Paired-associate (PA) transfer effects were compared for the A-B, C-A and A-B, B-C paradigms, using infrequently printed English words to constitute the PA lists and a mixed-list control (A-B, C-D paradigm) for nonspecific transfer. Negative transfer was obtained under both paradigms, with the A-B, C-A paradigm resulting in significantly greater negative transfer ($p < .01$) than the A-B, B-C paradigm. The results were in the same direction as those obtained by Murdock who found comparable (nonsignificant) negative transfer effects for the 2 paradigms.

Harcum (1953), as well as Murdock (1958), investigated paired-associate (PA) transfer effects under the A-B, C-A and A-B, B-C paradigms. Although negative transfer was obtained for both paradigms, Harcum found greater negative transfer for the A-B, B-C paradigm (as compared to the A-B, C-A paradigm) while Murdock obtained opposite

---

[1] Portions of this research were supported by United States Public Health Service Grant MH 06957-01, under the direction of the second author.

[2] Now at the Institute of Human Learning, University of California, Berkeley.

results; i.e., greater negative transfer for the A-B, C-A paradigm.

The present experiment had two purposes: (*a*) to replicate Murdock's experiment which used an appropriate control (A-B, C-D paradigm) for nonspecific transfer effects, using words (rather than nonsense syllables) to constitute the PA lists, the conventional 2:2-sec. rate of PA presentation, and a mixed-list control to increase the precision of measurement; and (*b*) to evaluate two different hypotheses regarding the magnitude of transfer to be expected between the paradigms.

The first hypothesis concerns competition of List 1 backward (R-S) associations with List 2 forward (S-R) associations. The A-B, C-A and A-B, B-C paradigms form B-A, C-A and B-A, B-C relationships, respectively, when evaluated in terms of List 1 R-S and List 2 S-R associations. The B-A, B-C paradigm (or in conventional transfer terminology, A-B, A-C) uniformly leads to pronounced negative transfer, while the B-A, C-A relationship (conventionally, A-B, C-B) usually results in slight negative transfer or, in other instances, positive transfer. Thus, with the competition hypothesis, the A-B, B-C paradigm should show greater negative transfer than the A-B, C-A paradigm. Kausler and Kanoti (1963) have recently demonstrated that List 1 R-S associations can compete or interfere with List 2 acquisition under the A-B, C-B paradigm, thus providing indirect support for this hypothesis.

The second hypothesis is related to stimulus familiarization (Schulz & Tucker, 1962). These authors have found that stimulus learning by overt articulation enhances subsequent PA acquisition, and, in the A-B, B-C paradigm, the B (List 2) stimuli have to be articulated and completely learned on List 1 in that they serve as response terms on this list. In the A-B, C-A paradigm, no List 2 stimulus familiarization is possible. Thus, this hypothesis would generate the opposite prediction of the first; i.e., greater negative transfer for the A-B, C-A paradigm than for the A-B, B-C paradigm.

Method.—A 2 × 2 "mixed" factorial design was used with transfer paradigms (A-B, C-A and A-B, B-C) as a "between" factor and a mixed-list control (A-B, C-D paradigm) as the "within" factor. Twenty general psychology students were randomly assigned to each of the "between" treatments and learned two consecutive PA lists consisting of four experimental and four control pairs. The PA pairs consisted of infrequently printed English words previously used by Horton and Kjeldergaard (1961). List construction was manipulated such that all pairs in List 2 were used equally often in the experimental or control conditions and such that one half of the *S*s assigned to each between treatment learned the same List 2. The lists were presented on a Lafayette memory drum (three random orders) at a 2:2-sec. rate with a 6-sec. intertrial interval. Learning was carried to a criterion of three consecutive errorless trials for each list.

Results and discussion.—List 2 performance was analyzed with respect to the number of correct anticipations on Trials 1–5 and also to the criterion of three perfect trials. Table 1 presents the mean correct anticipations and the mean difference scores (E − C) for each of the paradigms. The 2 × 2 analysis involving correct anticipations revealed the experimental vs. control (E/C) main effect to be significant for both analyses, *F*s (1, 38) = 17.15 and 22.83, $p < .001$, respectively, for the early and late criteria.

Negative transfer occurred under both transfer paradigms; i.e., performance on the control lists (A-B, C-D) was better. The paradigms main effect did not approach significance (both $Fs < 1$), nor was there a Paradigms × E/C interaction in either analysis, $F\ (1, 38) < 1$; $F\ (1, 38) = 2.07$.

Finally, analyses were performed to compare the transfer effects between the A-B, C-A and A-B, B-C paradigms. Measures of transfer were obtained by subtracting A-B, C-D performance (correct anticipations) from the experimental (transfer) paradigm performance for each *S*. Separate *t* tests involving the mean difference scores (E − C) were performed comparing the A-B, C-A and A-B, B-C transfer effects at each List 2 criterion. In each case, greater negative

TABLE 1

MEAN CORRECT ANTICIPATIONS ON LIST 2 ACROSS TRIALS 1–5 AND TO COMPLETE LEARNING

	A-B, C-A	A-B, C-D	E − C	A-B, B-C	A-B, C-D	E − C
Trials 1–5	7.00	9.70	−2.70	7.75	9.65	−1.90
Complete Learning	31.45	38.25	−6.80	34.75	38.00	−3.65

transfer was observed under the A-B, C-A paradigm. The five-trial analysis yielded a $t\ (38) = 1.45$, $p < .20$ ($s_{\bar{x}} = .55$), while the t at the late criterion was highly significant ($t = 3.09$, $p < .01$, $s_{\bar{x}} = 1.02$).

The comparisons of transfer effects between the A-B, C-A paradigms support the stimulus-familiarization hypothesis offered by Schulz and Tucker. It may be argued, however, that another process may account for the present results. This process involves the competition of List 1 S-R with List 2 R-S associations. When evaluated in this manner, the A-B, C-A and A-B, B-C paradigms yield A-B, A-C and A-B, C-B relationships in terms of List 1 S-R and List 2 associations, and, for the reasons discussed earlier in the paper, less negative transfer would be expected for the A-B, B-C paradigm.

REFERENCES

Harcum, H. R. Verbal transfer of overlearned forward and backward associations. *Amer. J. Psychol.*, 1953, **66**, 622–625.

Horton, D. L., & Kjeldergaard, P. M. An experimental analysis of associative factors in mediated generalization. *Psychol. Monogr.*, 1961, **75**(11, Whole No. 515).

Kausler, D. H., & Kanoti, G. A. R-S learning and negative transfer effects with a mixed list. *J. exp. Psychol.*, 1963, **65**, 201–205.

Murdock, B. B. "Backward" associations in transfer and learning. *J. exp. Psychol.*, 1958, **55**, 111–114.

Schulz, R. W., & Tucker, I. F. Supplementary report: Stimulus familiarization in paired-associate learning. *J. exp. Psychol.*, 1962, **64**, 549–550.

(Received June 20, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 5, 539–541

DIFFERENTIAL ADAPTATION OF ANXIOUS AND NONANXIOUS CHILDRF IN INSTRUMENTAL ESCAPE CONDITIONING[1]

RONALD K. PENNEY AND PAUL M. KIRWIN
University of Waterloo *Austin State School*

A group of 24 anxious children and a group of 22 nonanxious children (determined by Sarason's anxiety scales) were administered 20 tones of 84 db. each from which they could escape by pulling a lever. ½ of each group received adaptation training prior to this escape conditioning, i.e., were exposed to 15 tones of 5 sec. duration of 84 db. each. The other ½ of each group were not exposed to these tones (nonadapted groups). At the beginning of escape conditioning, the adapted group of nonanxious Ss exhibited slower mean lever response times than the nonadapted, nonanxious group. There were no differences in mean lever response times between adapted and nonadapted groups of anxious Ss.

The differential-adaptation hypothesis (Spence, 1956) assumes that high-anxious *S*s adapt at a relatively slower rate to situations containing a threat or some degree of noxiousness in comparison to low-anxious *S*s. Assuming the strength of the emotional response (r_e) is a determinant of drive strength and is reduced through adaptation training, the implication is that both high- and low-anxious *S*s would experience a decrement in drive as a result of adaptation, but the high-anxious group would show less of a decrement than the low-anxious group.

In applying this hypothesis to aversive conditioning situations, greater differences in the level of conditioning between high- and low-anxious groups would be expected when adaptation training preceded conditioning. That is, while the level of conditioning would be lower after adaptation for both groups, the difference between the groups would be greater after adaptation training with the high-anxious group showing less of a decrement in performance than the low-anxious group.

MacDonald (1946) and Taylor (1956) have reported an adaptation effect in finger-retraction and eyelid-conditioning situations. However, to date the interaction between adaptation and anxiety as suggested by the differential-adaptation hypothesis has not been investigated.

Method.—The lever apparatus consisted of a wooden box, which was painted gray. The facing panel sloped at an angle of 48° to the base. A lever, painted black, protruded through a slot in the facing panel and could be pulled downward and toward *S* through an excursion of 11 in. A coil spring returned the lever to its starting position at the end of each trial. A hand-shaped, white, wooden pattern was located at the bottom right, lower edge of the facing panel. The hand pattern was placed over a plunger type microswitch such that when *S* removed his hand from the pattern, a Standard Electric timer was activated. The timer was deactivated at the completion of the excursion of the lever. A tone generator was used in conjunction with two speakers to produce a 2,000-cps tone of 84 db. as measured by a General Radio sound-level meter held at a distance from the speakers at which maximum sound intensity was registered.

---

[1] The preparation of this report was facilitated by National Institute of Mental Health Grant M-4240-03 to the first author.

[2] The authors wish to express their thanks to R. M. Bryant, Principal of Walnut Creek Elementary School for his cooperation in support of this research.

Sarason's (1960) General Anxiety Scale for Children (GASC) and Test Anxiety Scale for Children (TASC) were administered to 170 children in Grades 4 and 5. The GASC and TASC scores for each *S* were combined and 24 *Ss* who had GASC scores ranging from 19 to 24, TASC scores from 11 to 22, and combined GASC and TASC scores from 35 to 59 were designated as a high-anxious group. The low-anxious group was composed of 22 *Ss* who had GASC scores ranging from 3 to 18, TASC scores from 0 to 12, and summed GASC and TASC scores from 6 to 30.

Half of the high-anxious *Ss* and half of the low-anxious *Ss* (adaptation groups) were given 15 5-sec. exposures to an 84-db. tone to permit adaptation to the tone prior to escape-conditioning trials. The *Ss* in these adaptation groups were individually seated facing the lever apparatus and were instructed to sit quietly and listen to the sounds that would follow. During the experiment, *E* remained behind a wooden screen to the right of *S*.

The remaining high- and low-anxious *Ss* were not preexposed to the tone prior to escape conditioning (nonadaptation groups). For the escape-conditioning phase, *S* was instructed to depress the white, wooden hand on the face of the apparatus until a sound was emitted after which he could depress the lever all the way to the bottom with his right hand. Depression of the lever through its full excursion terminated the tone.

During escape conditioning, 20 trials of the 84-db. tone were administered to all *Ss*. The time interval between trials was not constant but varied with the maximum intertrial interval of 10 sec.

Results and discussion.—The mean response times for the four groups were arranged in five blocks of four trials each. An analysis of variance of these data revealed a significant Trials × Anxiety × Adaptation triple interaction, $F\ (4, 168) = 15.50$, $p < .001$, indicating that the difference between adapted and nonadapted groups was a function of trials and anxiety. Accordingly, two additional analyses were conducted. The first analysis compared the adapted and nonadapted groups of high-anxious *Ss* over the five blocks of trials. This analysis yielded a significant Trials effect, $F\ (4, 88) = 34.00$, $p < .001$. The main effect for Adaptation and the interaction between Adaptation and Trials were not significant. The second analysis compared the adapted and nonadapted groups of low-anxious *Ss* over the five blocks of trials. The Trials effect, $F\ (4, 80) = 36.75$, $p < .001$, and the Trials × Adaptation interaction, $F\ (4, 80) = 4.75$, $p < .005$ were significant. This significant interaction indicated that the low-anxious adapted group exhibited slower mean response times than the low-anxious, nonadapted group as a function of trials. When Cochran's approximate t test was applied to the five blocks of trials, it was found that for low-anxious *Ss*, the adapted group was slower in its mean response time over the first block of trials relative to the nonadapted group ($t = 2.43$, $p < .05$). These two groups did not differ at any of the other blocks of trials.

Figure 1 presents the mean total response times of the four groups for the first block of trials. A 2 × 2 factorial analysis of variance was applied to these data. The interaction between adaptation and anxiety was significant, $F\ (1, 42) = 4.19$, $p < .05$. In testing for the simple effects, the only significant

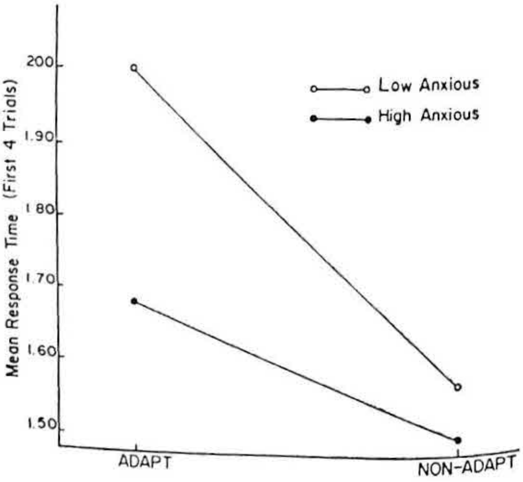


Fig. 1. Mean response times as a function of adaptation (adapt vs. nonadapt) and anxiety (high vs. low).

difference found was between the adapted and nonadapted groups of low-anxious *Ss*, $t\ (20) = 2.11$, $p < .05$. The F for adaptation just failed to reach significance, $F\ (1, 42) = 3.84$, $p < .07$.

The differential adaptation of high- and low-anxious *Ss* appears to be supported by the results of the present study. Whereas the adapted group of low-anxious *Ss* conditioned more slowly than the nonadapted group at the beginning of training, no differences between adapted and nonadapted groups were found for high-anxious *Ss*. A question arises as to why the adapted group of low-anxious *Ss* did not continue to condition more slowly beyond the first block of trials. Since the present study is one of the few studies to investigate the relationship between anxiety and adaptation, any explanation offered to account for this finding must be speculative. One explanation involves the assumption that the relationship between the inferred emotional response (r_e) and number of adaptation trials is curvilinear. That is, after the presentation of "*x*" noxious stimuli, r_e is gradually increased in strength again. It seems reasonable to assume that the presentation of noxious stimuli may at first reduce r_e but beyond a certain number of trials, the accumulation of noxious stimuli may heighten r_e due to the large number of previous presentations. Further investigations will be needed to determine the relationship between number of adaptation trials and performance in aversive-conditioning situations before inferences can be made concerning the relationship of r_e to number of adaptation trials.

REFERENCES

MacDonald, A. The effect of adaptation to the unconditioned stimulus upon the formation of conditioned avoidance responses. *J. exp. Psychol.*, 1946, **36**, 1-12.

Sarason, S., Davidson, K., Lighthall, F., Waite, R., & Ruebush, B. *Anxiety in elementary school children.* New York: Wiley, 1960.

Spence, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

Taylor, J. A. Level of conditioning and intensity of the adaptation stimulus. *J. exp. Psychol.*, 1956, **51**, 127-130.

(Received July 6, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 5, 542

Beginning with the first issue of Volume 69 and continuing through the remainder of that volume and Volume 70 (1965), the titles and authors of accepted papers will be listed here following Supplementary Reports. It is being supported on an experimental basis by the APA Project on Scientific Information Exchange in Psychology, and at the end of the year, the outcome of this trial will be evaluated and consideration given the advisability of continuing the listing.

This listing plus those published in the preceding 1965 issues are a portion of the entire backlog of manuscripts accepted by this Journal. Such listing will allow readers to become aware of research many months in advance of journal publication.

Manuscripts Accepted for Publication in the *Journal of Experimental Psychology*

Human Heart-Rate Responses during Experimentally Induced Anxiety: Effects of Instructions on Acquisition: George E. Deane*: Department of Psychology, Harpur College, State University of New York, Binghamton, New York.

Spatial Adaptation and Aftereffect with Optically Transformed Vision: Effects of Active and Passive Responding and the Relationship Between Test and Exposure Responses: G. Singer and R. H. Day*: Department of Psychology, Monash University, P.O. Box 92, Clayton, Victoria, Australia.

Inference Behavior in Multiple-Cue Tasks Involving Both Linear and Nonlinear Relations: David A. Summers* and Kenneth R. Hammond: Institute of Behavioral Science, University of Colorado, Boulder, Colorado 80304.

Resistance to Extinction when Continuous Reinforcement is Followed by Partial Reinforcement: David Hothersall*: Department of Psychology, Birkbeck College, Malet Street, London W.C.1, England.

Effects of CS and UCS Relationships on Electrodermal Response and Heart Rate: George H. Zimny,* John A. Stern, and Stanton P. Field: Department of Neurology and Psychiatry, Saint Louis University, David P. Wohl Memorial Mental Health Institute, 1221 South Grand Boulevard, Saint Louis, Missouri 63104.

Preference and Recall of Stimulus Variability: Harry Munsinger* and William Kessen: Department of Psychology, Yale University, 333 Cedar Street, New Haven, Connecticut.

Temporal Factors in Verbal Satiation: Dorothy Hubbard Gampel*: Department of Psychology, The City College, The City University of New York, New York, New York 10031.

Adaptation to Prismatic Displacements: Hand Position and Target Location: Robert Wm. Sekuler* and Joseph A. Bauer, Jr.: Department of Psychology, Northwestern University, Evanston, Illinois.

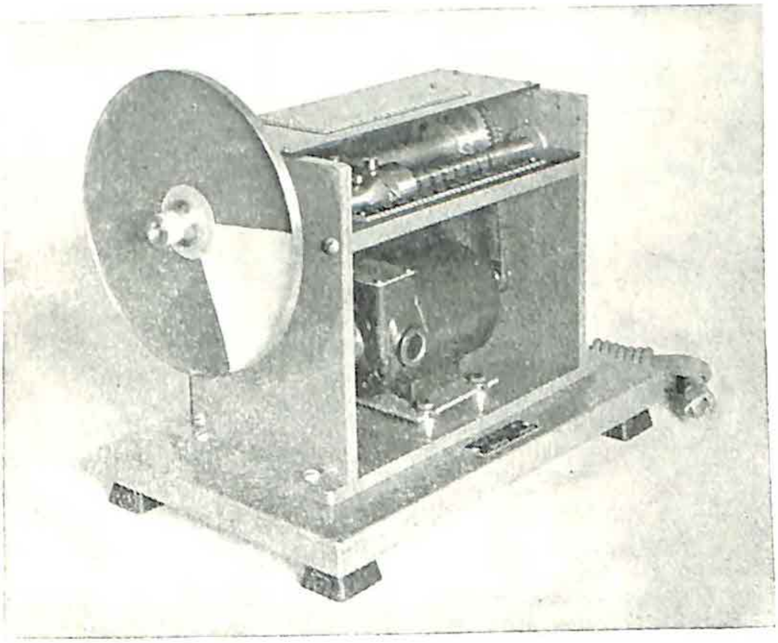
Eyelid Trace Conditioning, CS Intensity, CS-UCS Interval, and a Correction for "Spontaneous" Blinking: Stanley G. Lipkin and John W. Moore*: Department of Psychology, University of Massachusetts, Amherst, Massachusetts 01003.

Transfer of Verbal Discriminations Based on Differential Reward Magnitudes: W. K. Estes*: Department of Psychology, Stanford University, Stanford, California 94305.

---

* Asterisk indicates author for whom address is supplied.

DIFFERENTIAL COLOR MIXER



The construction of this color mixer permits the observer to vary the proportions of two color discs while they are being rotated. It is a machine of high precision which uses color discs made from standard color papers. Complicated fasteners for the discs are not required. The maximum diameter of the paper discs it will accommodate is 8″. The total angular displacement of the two discs is approximately 340 degrees. The scale which indicates this displacement is calibrated in 5 degree units. Settings on this scale can be estimated to one degree.

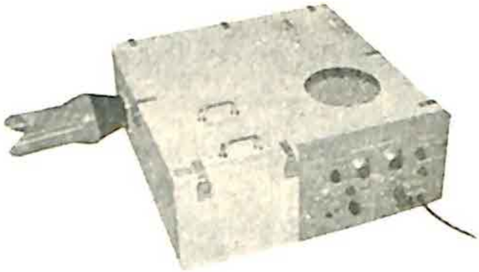
Ralph Gerbrands Company

Scientific Instruments

8 Beck Road
Arlington, Mass. 02174

Tel. 648-6415
Area Code 617

NEW . . . FROM SCIENTIFIC PROTOTYPE



A LOW COST · HIGH PERFORMANCE FULL CAPABILITY · TWO CHANNEL TACHISTOSCOPE

A new tool for the researcher. A two-channel Tachistoscope for use where all the extraordinary features of the world reknown 3-channel Scientific Prototype GB Tachistoscope's are not required. The 2-channel Model 800E, however, . . . at a far lower price . . . offers all the accuracy, proven features, and the quality that made the Model GB famous.

Much of the design and capability of the Model 800E is based on the suggestions and information provided by researchers and teachers in the field.

Thus the following is just a few of the areas that can be investigated:

BASIC VISUAL PSYCHOPHYSICS a. Area-time relationships (Ricco's and Piper's laws) b. Intensity-time relationships (Bloch's and Pieron's laws) c. Dark and Light Adaption d. Relative and Absolute Brightness Thresholds e. Visual Acuity f. Broca-Sulzer Phenomenon. PERCEPTION—a. Perceptual Defense b. Duration Thresholds for Symbols, Words, Numbers, etc. c. Figure-Ground Phenomena d. Contour Perception e. Figural After-Effects.

FEATURES: Accepts standard 4" × 5" stimulus cards or transparencies. ● Electronic system contains two separate time internal generators. ● Range 1.0 msec to 11 sec. in four overlapping direct reading ranges. ● Accuracy guaranteed ± 2% of any setting. ● Electronic intensity controls over approximately 1 log unit. ● Field locks allow direct operator intervention without disturbing logic. ● Logic provided to allow continuous alternating of fields, overlapping (competitive) exposures and the use of one field as a pre and post exposure adaption and fixation field, while other field is used for stimulus exposure. ● Blower cooled electronics for minimum temperature rise, maximum reliability. ● Jack provided for remote firing of unit. Complete logic control for each timer, plus inhibit logic, plus manual field lock switches, and a solid state pulse generator which may be controlled by remote switch closure, provide maximum flexibility. ● Two lamps per field—rise and fall times guaranteed under 0.01 milliseconds, laminated front surface mirror, with magnesium fluoride anti-reflection back coating. Adjustable mirror mount for field alignment, four internal masks with adjustable edges for stimulus framing and elimination of internal cues.

Only four vacuum tubes (lamp switches), all power supply, logic, lamp current control and timer semiconductors, are silicon for maximum reliability and negligible temperature dependence. Logic and lamp current reference power supplies are regulated. Timing and lamp current are independent of line voltage.

Electronic intensity controls plus filter holders for neutral density filters or transparent stimuli (4" × 5"). Reversible holders for 4" × 5" stimulus cards.

Special Features: High voltage interlock, Low noise forced ventilation, Fitted viewing hood, Compact self-contained construction. No inter-connecting cables required.

Priced $845, f.o.b., New York City. Delivery from stock.

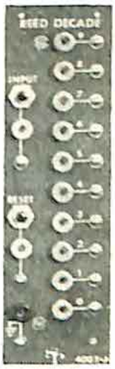
Lab Manual Available: Send $3 for "10 Experiments in Vision for 2-Channel Tachistoscope"

SCIENTIFIC PROTOTYPE MFG., CORP.

Behavior Analysis Apparatus . . . reed relay control equipment cages, pellet dispensers, soundproof chambers, recorders, noise generators, shock sources, timers, switches, digital test equipment and special systems.

615 W. 131st Street
New York, New York 10027
Telephone: (212) 368–6000

REED RELAY EQUIPMENT

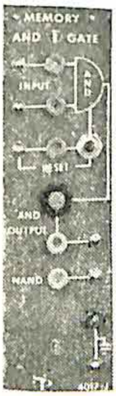


The Scientific Prototype Reed Relay System of operant programming control equipment represents a new approach to the design & construction of relay equipment which is fully compatible with and functionally equivalent to conventional 24 VDC equipment. The new system provides higher operating speed, lower noise, longer life, simpler interconnection, greater flexibility and improved reliability at a competitive price. Input and output isolation, separate input and output spark suppression and a host of internal refinements provide a new high in convenience and utility.

The individual units are designed to clip mount on standard 6½″ 24 to 28 VDC power rods. Each unit requires ⅛, ¼ or ½ the space provided on a single pair of bars. This system *must* be operated with the positive power supply lead attached to the top power rod of each pair. A "lock-out" on each unit will prevent operation if the polarity is reversed. Power supply voltage must be held to 24 to 28 VDC including ripple and transients. A momentary drop below 22 VDC may cause spurious "drop-out." A momentary excursion above 30 VDC may cause damage to semiconductors and will void the warranty. Scientific Prototype reed relay power supplies are well regulated and provide current limiting plus overload "lock-out" for double short circuit protection.

All inputs and outputs are in the conventional form of switch closures to ground (— 24 VDC) except special outputs such as shock, audio signals, etc. All outputs are rated at 3 amps except for the stepper decade and a 2-pole, 33-position stepper (which are Automatic Electric type 40 steppers rated at 300 milliamps interrupt, 1 amp carry), and the logic and cradle relays, binary relay and manual timer (rated at 5 amps). All solid state outputs are internally fused at 3 amps (make, break and carry, 24 to 28 VDC resistive or inductive). Inputs draw low current (30 to 100 milliamps typically) and are buffered by reed relays or transistors.

Two types of inputs are used: "A.C." (pulse coupled and "D.C." (direct coupled). An A.C. input responds to the onset or establishment of an input; the continued presence of an input has no effect. A D.C. input responds to the presence of an input signal. The response continues for the entire duration of the input. Similarly, A.C. and D.C. outputs are used. A.C. outputs provide a pulse to signal the establishment of a particular condition. The continued presence of the condition does not produce any further output. A D.C. output signals the presence of a particular condition and continues as long as the condition persists.



There are two front panel connections for each input and output. One member of each pair is a conventional snap stud, to insure compatibility with existing relay equipment, the other is a color coded pin jack which accepts stackable patch cords. While snap leads may be used throughout the system, pin plugs provide a more durable and reliable interconnection system and the pin jacks themselves give visual indication of the function of the connection point.

Operating speeds for reed relay units are conservatively given at 100 cps and 1000 cps for solid state units unless deliberately limited by filters as in a slo-make, slo-break relay, bounce filter, delay generator, etc. Units containing standard electro-mechanical devices such as Sodeco counters are limited by the operating characteristics of these devices.

SCIENTIFIC PROTOTYPE MFG., CORP.

Behavior Analysis Apparatus . . . reed relay control equipment cages, pellet dispensers, soundproof chambers, recorders, noise generators, shock sources, timers, switches, digital test equipment and special systems.

615 W. 131st Street **New York, New York 10027**

Telephone: (212) 368-6000

Journals of the

AMERICAN PSYCHOLOGICAL ASSOCIATION

Subscription Rates for 1966

American Psychologist. Official papers of the Association and articles on psychology. Monthly. The 1966 volume is Vol. 21. Subscription: $10.00 (Foreign $10.50). Single copy $1.00.

Contemporary Psychology. Critical reviews of books, films, and research material in the field of psychology. Monthly. The 1966 volume is Vol. 11. Subscription: $10.00 (Foreign $10.50). Single copy $1.00.

Journal of Abnormal Psychology. Theoretical and research contributions in behavior pathology. Bimonthly. The issues appear in February, April, June, August, October, and December. The 1966 volume is Vol. 71. Subscription: $10.00. (Foreign $10.50). Single copy $2.00. (Continues volume sequence of the former *Journal of Abnormal and Social Psychology*.)

Journal of Applied Psychology. Applications of psychology to business and industry. Bimonthly. The issues appear in February, April, June, August, October, and December. The 1966 volume is Vol. 50. Subscription: $10.00 (Foreign $10.50). Single copy $2.00.

Journal of Comparative and Physiological Psychology. Original contributions in the field of comparative and physiological psychology. Bimonthly, two volumes per year. There are three issues in each volume. The issues appear in February, April, June, August, October, and December. The 1966 volumes are Vols. 61 and 62. Subscription: $20.00 (Foreign $20.50). Single copy $4.00.

Journal of Consulting Psychology. Research in clinical psychology: psychological diagnoses, psychotherapy, personality, psychopathology. Bimonthly. The issues appear in February, April, June, August, October, and December. The 1966 volume is Vol. 30. Subscription: $10.00 (Foreign $10.50). Single copy $2.00.

Journal of Educational Psychology. Studies of learning and teaching: measurement of psychological development, psychology of school subjects, methods of instruction, school adjustment. Bimonthly. The issues appear in February, April, June, August, October, and December. The 1966 volume is Vol. 57. Subscription: $10.00 (Foreign $10.50). Single copy $2.00.

Journal of Experimental Psychology. Original contributions of an experimental character. Monthly, two volumes per year. There are six issues in each volume. The 1966 volumes are Vols. 71 and 72. Subscription: $20.00 for 2 vols. (Foreign $20.50). Single copy $2.00.

Journal of Personality and Social Psychology. Theoretical and research papers on personality dynamics, group process, and the psychological aspects of social structure. Monthly, two volumes per year. There are six issues in each volume. The 1966 volumes are Vols. 3 and 4. Subscription: $20.00 for 2 vols. (Foreign $20.50). Single copy $2.00.

Psychological Abstracts. Noncritical abstracts of the world's literature in psychology and related subjects. Monthly. The 1966 volume is Vol. 40. Subscription: $20.00 (Foreign $20.50). Single copy $2.00.

Psychological Bulletin. Evaluative reviews of research literature, discussions of research methodology in psychology. Monthly, two volumes per year. There are six issues in each volume. The 1966 volumes are Vols. 65 and 66. Subscription: $20.00 (Foreign $20.50). Single copy $2.00.

Psychological Monographs: General and Applied. Longer researches and laboratory studies which appear as units. Published at irregular intervals during the calendar year: between 15 and 20 issues per year. The 1966 volume is Vol. 80. Subscription: $10.00 (Foreign $10.50). Single copy $1.00 to $3.00, according to size.

Psychological Review. Original contributions of a theoretical nature. Bimonthly. The issues appear in January, March, May, July, September, and November. The 1966 volume is Vol. 73. Subscription: $10.00 (Foreign $10.50). Single copy $2.00.

Cover pages, table of contents, and index are contained in the last issue of the volume for each journal.

Discounts: AGENTS and BOOK DEALERS receive a 10% discount on all orders.
There is no club arrangement.
Note: There are no quantity, cash, library, or other special discounts.

▶ SUBSCRIPTIONS ARE AVAILABLE ON A JANUARY THROUGH DECEMBER BASIS ONLY ◀

Payment must be made in U. S. funds.
Send subscription orders to:
AMERICAN PSYCHOLOGICAL ASSOCIATION
Subscription Department
1200 Seventeenth Street, N.W.
Washington, D. C. 20036

Journal of
Experimental Psychology

Vol. 70, No. 6 December 1965

DISCRIMINATED AVOIDANCE LEARNING AS A FUNCTION OF PARAMETERS OF DISCONTINUOUS SHOCK[1]

M. R. D'AMATO

Rutgers University

DONALD KELLER and GERALD BIEDERMAN

New York University

The parameters of discontinuous shock, which has been shown to facilitate markedly discriminated bar-press avoidance learning in the rat, were investigated in a study in which 3 values of on time (.2, .5, and 1.0 sec.) were combined factorially with 3 values of off time (.5, 1.0, and 2.0 sec.). It was concluded from the resulting data, as well as from other results, that the optimal values of discontinuous shock were 2 sec. off time and as brief an on time as would guarantee receipt of shock during every on period. Percentage of escape responses executed during off periods correlated highly with avoidance performance (rho = .864); escape latency correlated less well (rho = .733). The facilitory effect of discontinuous shock was ascribed primarily to enhanced generalization of the escape response to the preshock period.

The inordinate difficulty often encountered in obtaining discriminated avoidance learning in the bar press and related situations has posed practical and theoretical problems (e.g., D'Amato & Schiff, 1964; Meyer, Cho, & Wesemann, 1960). On the one hand, the inability to obtain with rats consistent avoidance conditioning in these experimental situations has deprived investigators of an important research tool. On the theoretical side, there seems to be no satisfactory way of accounting for the failure to learn a task which appears to be well within the behavioral capabilities of rats.

Difficulty in acquiring a discriminated avoidance response is by no means restricted to the rat level. The failure rate in dogs can be quite high (e.g., Brush, 1957); and as Turner and Solomon (1962) have shown, failures to learn a discriminated avoidance task occur under certain circumstances even with human *S*s.

Recently we showed that discriminated avoidance learning could be facilitated markedly by employing as the unconditioned stimulus (US) a

---

[1] This research was supported by National Science Foundation Grant GB-1515.

discontinuous, rather than the conventional continuous shock, i.e., a shock having brief on times and longer off periods (D'Amato, Keller, & DiCara, 1964). The hypothesis that led to the use of discontinuous shock was that *S*'s failure to learn to avoid was due in part to inadequate generalization of the escape response to the preshock period. Under usual training procedures the escape response occurs in the presence of shock, which probably comprises the major stimulus component acting when the instrumental response is executed. If the escape response could be made to occur in the absence of shock—as is the case when *S* responds during an off period of discontinuous shock—generalization to the preshock period should be facilitated, a conjecture borne out by our results.

The major purpose of the present study was to investigate the effects on avoidance performance of varying systematically the duration of on periods and off periods. We also wished to determine whether, in accordance with our hypothesis, a close relationship would be obtained between the extent of the facilitating effects of discontinuous shock and the percentage of escape responses executed in off periods.

METHOD

Subjects and apparatus.—The *S*s were 102 experimentally naive albino rats (50 males) drawn from litters bred in our laboratory; they were about 70–145 days of age at the start of the study. The apparatus consisted of two commercial Skinner boxes and associated programing equipment, described in a previous report (D'Amato et al., 1964).

Design.—Three on times, .2, .5, and 1.0 sec., were combined factorially with three off times, .5, 1.0, and 2.0 sec. In addition a group of 12 *S*s was run with continuous shock (0 sec. off time). Originally there were 8 *S*s in each of the nine discontinuous shock groups, but in order to obtain additional data for the three .5-sec. off-time conditions, a replication consisting of 6 *S*s in each of the three cells was run. Since there were no important differences between replications the data were pooled, providing 14 *S*s in the latter three groups.

Procedure. On Day 1 of training 40 avoidance trials were given, continuous shock being employed in all groups. During these trials *S*s were shaped on the escape response until they acquired a reasonable degree of proficiency. The shock, delivered through a Grason-Stadler (Model E1064) grid-scrambling device, was usually .6 ma. for females and .8 ma. for males.[2] The CS was an increase in illumination level—from .5 to 3.1 footcandles (ftc.)—provided by powering the three pilot lights on the front wall of the experimental chamber. The CS-US interval was 5 sec., and the intertrial interval, programed by a film tape, averaged about 35 sec. When *S* failed to avoid, i.e., failed to respond during the CS-US interval, the US remained on until *S* made the escape response, which terminated CS and US.

On Day 2, 500 trials were administered under the appropriate US condition. With the onset of the US a pulse of shock corresponding in duration with the on-time assignment was delivered, followed by the appropriate off period, the cycle being repeated until *S* made the escape response, terminating both CS and US. Shock was increased to .8 ma. for females and 1.0 ma. for males. Escape responses occurring during shock delivery or approximately .25 sec. after the termination of shock were classified as occurring during on periods; escape responses not in this interval were recorded as having fallen in off periods. Escape and avoidance responses were recorded on counters in blocks of 25 trials.

RESULTS AND DISCUSSION

The major dependent variable used in our analyses was the percentage of avoidances achieved over the last 400 trials of Day 2, a measure employed in previous studies. In Fig. 1 group means based on this measure are shown for the 10 groups of the experiment. It is clear that with continuous shock (0 sec. off time) only a very low level of avoidance responding is attained. An off period of .5 sec. is a little better than 0 off time, but the difference between the respective means (26.7% vs. 17.9%) is far from significant, $t = .89$. As off time is increased to 1.0 and 2.0 sec., avoidance performance also increases, and at these off periods the potent effect of on time becomes apparent.

---

[2] The shock intensities reported refer to dial settings on the shocker and probably are a little higher than the actual values (cf. Annau & Kamin, 1961).

Analysis of variance showed off time to be a significant factor, $F(2, 81) = 3.79$, $p < .05$; neither on time nor the interaction reached conventional significance levels, $F(2, 81) = .79$ and $F(4, 81) = 1.31$, respectively. The failure of on time to emerge as a significant factor is largely due to the ineffectiveness of this variable at the .5 sec. off time. With this value of off time removed from the analysis, on time comes quite close to the .05 level of significance, $F(2, 81) = 3.04$.

Two conclusions seem warranted from the data of Fig. 1. (*a*) With off periods sufficiently long for on time to be an effective variable, the shorter the on time the higher the level of avoidance learning. In all likelihood on times shorter than .2 sec. would facilitate avoidance learning further, and this impression is consistent with data obtained by us with .15-sec. on periods, but there is a limit to which on time can be reduced while still guaranteeing that *S*s receive shock during each on period. (*b*) Off times as short as .5 sec. are little, if at all, better than continuous shock, regardless of the on time employed.

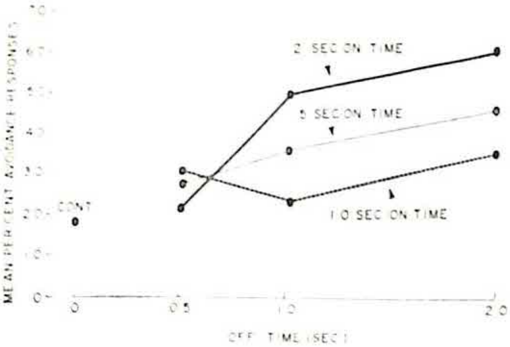


FIG. 1. Mean percentages of avoidance responses over the last 400 trials of Day 2 as a function of shock on and off times. (Continuous shock is represented as 0 sec. off time.)

Since facilitation of avoidance performance increases throughout the range of off times employed in this experiment, the question arises whether additional enhancement would occur with off periods longer than 2 sec. The evidence available to us seems to indicate not. Prior to the present experiment we ran a study comparing 1-, 2-, and 4-sec. off periods, on time being held constant at .15 sec. The only other procedural difference between the two studies was that a white noise, rather than a visual cue, served as the CS. The mean percentages of avoidances over the last 400 trials were, in order, 69.0, 75.9, and 63.5 ($N = 8$ per group). Though the differences between the groups are not significant, they suggest that something in the order of 2 sec. might be the optimal off time. This conclusion is supported further by the fact that a trend analysis over the first 200 trials of Day 2 in blocks of 50 trials showed the linear component of the 2-sec. off time group to be steeper than that of the 4-sec. group ($p = .05$).

On the basis of the above results, as well as from indications arising from other data gathered by us, we conclude provisionally that the optimal parameters of discontinuous shock are 2 sec. off time and the briefest possible on time that still assures receipt of shock during each on period.

Escape responses in off periods vs. latency as predictors of avoidance performance.—If, as we have maintained, discontinuous shock facilitates avoidance learning because escape responses, often falling into off periods, are more likely to generalize to the preshock interval, there should be a close correspondence between avoidance performance and percentage of escape responses executed during off times. Consequently, for each *S* we calculated over the first 100 trials of Day 2 the percentage of all escape responses that occurred during off periods. Actually, because we wished to exclude at least some of the responses that were initiated during shock, we required that for an escape response to be classified as falling in an off period, it must occur no sooner than approximately .25 sec. after the termination of an on period. We chose the first 100 trials because this is the portion of training when the differences between the different USs seem to take hold and yet even in the best conditions *S*s are still making a fair number of escape responses. The group means based on this measure are plotted in Fig. 2, and it is apparent that particularly for the 1.0- and 2.0-sec. off-time groups there is a good correspondence between the pattern displayed here and that of Fig. 1. A rank-order correlation based on all 10 groups yields a rho of .864 ($t = 4.85$, $p < .002$).

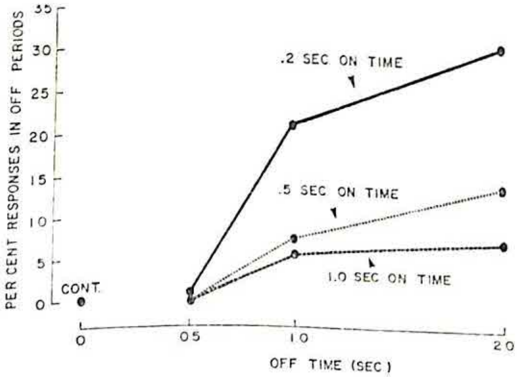


FIG. 2. Mean percentages of escape responses in off periods over the first 100 trials of Day 2 as a function of shock on and off times, with continuous shock represented as 0 sec. off time.

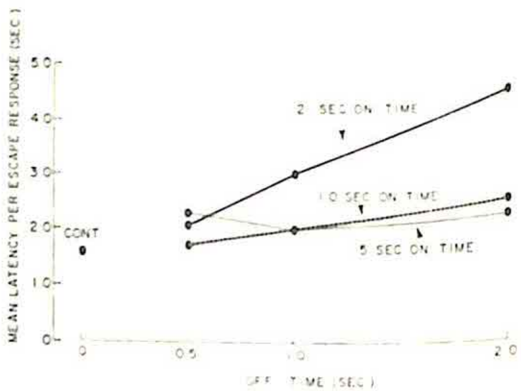


FIG. 3. Mean latency per escape response over the first 100 trials of Day 2 as a function of shock on and off times. (Continuous shock is represented as 0 sec. off time.)

An alternative explanation of the facilitating effects of discontinuous shock may be derived from the notion of response "reflexiveness" advanced by Turner and Solomon (1962), one of the very few formulations specifically designed to account for failures of avoidance learning as well as for successes. Very briefly their view is that the more reflexive a response chosen for avoidance conditioning, the poorer the avoidance learning. One important criterion of reflexiveness is response latency. The "same" response environmentally defined—e.g., bar pressing—can differ in reflexiveness if its latency is by some means manipulated, as, e.g., by varying the intensity of the US (Turner & Solomon, 1962, p. 2). If it is permissible to apply their theory to the present situation, the degree of reflexiveness of the instrumental response will vary from one US condition to another since, in general, escape latency will be a function of on and off times.

In order to evaluate this approach, we calculated for each *S* the average latency per escape response for the first 100 trials of Day 2. The 10 group means based on this measure are presented in Fig. 3. There is a general correspondence between these scores and avoidance behavior (Fig. 1), though perhaps not as close as was observed for responses in off periods. Rho turns out in the present case to be .733 ($t = 3.05$, $p < .02$), so that latency is not quite as good a predictor of avoidance performance as escapes in off periods.

Further evidence that responses in off periods is the better predictor of avoidance performance is provided by the results of the study in which off time was carried to 4 sec. It will be recalled that the 2-sec. off-time group averaged more avoidance responses over the last 400 trials than the 4-sec. off-time group, 75.9% vs. 63.5%. Although an added opportunity for responding in off periods existed for the 4-sec. group, it turns out the 2-sec. off-time group averaged the greater number of escape responses in off periods, 61% vs. 48%. In terms of the latency measure, however, the 2-sec. off-time group had a *lower* average latency than the 4-sec. group, 5.16 vs. 5.57 sec.

Taken together, then, the present results lend support to our view that the facilitating effect of discontinuous shock is attributable to enhanced generalization of the escape response to the preshock period. As to the means by which the increased generalization is mediated, two different mechanisms seem to be involved. The first, and probably the most important of the two, has to do with the instrumental response itself. With optimal discontinuous shock the escape response quickly develops into a well-organized response sequence, closely resembling the avoidance behavior to which it eventually gives rise. With continuous shock, on the other hand, even at the moderate shock levels employed in our experiments the escape response often takes the form of a quick, sometimes frenzied, lunging at the bar, a response sequence quite unlikely to generalize directly to the CS alone. Sometimes such escape responses do generalize to the preshock period, and it is a compelling sight to see an animal, duplicating his response to the US, leap toward the bar at the onset of the CS. But such avoidance behavior is effortful, inefficient, and unstable; and *S* either soon stops avoiding or develops a more efficient avoidance response.

The second mechanism, more motivational in nature, is conditioned suppression—the evocation in *S* by the CS (owing to its association with shock) of a posture of crouching and freezing that effectively inhibits a wide range of instrumental behavior. Other things equal, conditioned suppression very probably develops to a much greater degree when continuous, rather than discontinuous, shock serves as the US. The basis of this presumed difference might simply reside in a difference between the aversiveness of the two types of shock, a possibility which fits in with the fact that the conditioned emotional response has been found to be directly related to shock intensity (Annau & Kamin, 1961). Though we do not as yet have direct support for the preceding hypothesis, evidence of an indirect nature exists (Biederman, D'Amato, & Keller, 1964).

Whatever the bases of their effectiveness turn out to be, there can be no doubt that the parameters of discontinuous shock comprise an extremely potent pair of relevant variables in the discriminated bar-press avoidance situation.

REFERENCES

ANNAU, Z., & KAMIN, L. J. The conditioned emotional response as a function of intensity of the US. *J. comp. physiol. Psychol.*, 1961, **54**, 428-432.

BIEDERMAN, G. B., D'AMATO, M. R., & KELLER, D. M. Facilitation of discriminated avoidance learning by dissociation of CS and manipulandum. *Psychon. Sci.*, 1964, **1**, 229–230.

Brush, F. R. The effects of shock intensity on the acquisition and extinction of an avoidance response in dogs. *J. comp. physiol. Psychol.*, 1957, **50**, 547–552.

D'Amato, M. R., Keller, D. M., & DiCara, L. Facilitation of discriminated avoidance learning by discontinuous shock. *J. comp. physiol. Psychol.*, 1964, **58**, 344–349.

D'Amato, M. R., & Schiff, D. Long-term discriminated avoidance performance in the rat. *J. comp. physiol. Psychol.*, 1964, **57**, 123–126.

Meyer, D. R., Cho, C., & Wesemann, A. F. On problems of conditioning discriminated lever-press avoidance responses. *Psychol. Rev.*, 1960, **67**, 224–228.

Turner, L. H., & Solomon, R. L. Human traumatic avoidance learning: Theory and experiments on the operant-respondent distinction and failures to learn. *Psychol. Monogr.*, 1962, **76**(40, Whole No. 559).

(Received July 28, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 6, 549-558

EFFECTS OF DISTRIBUTION OF PRACTICE OF THE ACQUISITION PAIRS UPON MEDIATING RESPONSES AND RESPONSE AVAILABILITY[1]

MARGARET JEAN PETERSON AND SARAH M. JAMISON

Indiana University

Distribution of practice of the acquisition pairs in a mediation paradigm was manipulated by inserting 0, 2, or 5 items between the 3 presentations of the acquisition pairs, B-C and B-A. Some evidence of mediated facilitation was shown with massed but not with distributed practice when less stringent tests of retention of the acquisition pairs were used. Mediated facilitation was found for all levels of distribution with more stringent retention tests. Response availability (difference between empirical levels of control responding and a chance expectation) was inversely related to distribution of practice of the acquisition pairs. Errors were complexly related to the manipulated variables.

Underwood and Schulz (1961) concluded that distribution of practice may facilitate learning when the response-learning phase involves some interference effects. Generalizing from their conclusion, distributed practice might influence mediation if the mediation paradigm incorporated an interference design. Goss (1961) and Jenkins (1963) suggested that some mediation paradigms produce interference which may tend to reduce mediational effects. One such paradigm, a response-equivalence type, B-C, B-A, A-C, actually yielded the greatest evidence of mediation in a study designed to compare a number of mediation paradigms (Peterson, Colavita, Sheahan, & Blattner, 1964)! The present experiments were conducted to explore the possibility that distribution of practice might influence mediation when the interference-type design, B-C, B-A, A-C, was used.

Distribution of practice was manipulated by varying the separation of three presentations of the first acquisition pair, B-C, and of the three presentations of the second acquisition pair, B-A. The same procedure was followed with the first two stages of the control paradigm, X-C and B-A. The effect of distributed practice upon mediated facilitation was assessed by comparing the performance on the test trials following mediation presentation, A-C, with the performance following control presentation, also A-C. The role of response availability was inferred from the degree to which the selections of the critical response on the test trials following control presentation differed from a chance expectation.

The paradigm used in Exp. I incorporated "cue trials" to test the immediate retention of the B-C and the B-A acquisition pairs. A cue trial presented the stimulus alone immediately after the stimulus-response pair—a "prompting" procedure. It was possible, under the conditions of Exp. I, that the associations necessary for a mediating response were not retained for the intervals

---

[1] These investigations were supported by Public Health Service Research Grant MH 07944, from the National Institute of Mental Health.

between the separated repetitions of the acquisition pairs with distributed practice and hence were not available to mediate behavior on the test trials. To provide a more stringent test of the retention of the acquisition pairs the cue-trial procedure was reversed in Exp. II so that the stimulus alone preceded the stimulus-response pair—the more usual anticipation technique. The reversed order was used for the second and third repetitions of the acquisition pairs. The initial presentation of the pair was followed by the cue trial, as in Exp. I.

METHOD

Learning materials.—Consonant-vowel-consonant trigrams, CVCs, of 87–100% association value (Underwood & Schulz, 1960) were used to construct mediation sets. A mediation set began with the first-stage presentation of a pair of CVCs, B-C. Stimulus B was then shown again as a cue trial to test retention of the previously presented pair and *S* was instructed to read off the stimulus and to give the response which had been paired with it on the previous presentation. The first acquisition pair and its cue trial were presented three times before the second acquisition pair was shown. The second acquisition pair consisted of the stimulus from the first pair, B, and another CVC response term, A. The cue trial, B, followed. The second acquisition pair and its cue trial were shown three times before the test trial was presented. The test trial was always shown immediately after the third presentation of the second acquisition pair and its cue trial. In the test trial, the response term from the second pair, A, was presented as the stimulus for the response term from the first pair, C, and two additional response alternatives, D and E. The *Ss* were instructed to verbally pair each of the response trigrams successively with the stimulus trigram A and then to select the response alternative (i.e., C, D, or E) which made the "best pair" with Stimulus A. A control set was prepared for each mediation set by substituting a new extraneous trigram for Stimulus B in the first stage pair, X-C, providing a pool of 36 mediation and their 36 control sets.

The C trigram appeared equally often in the first, second, or third position among the response alternatives on the test trials. To equate frequency of experience with each of the test-trial alternatives over the entire experiment, the D and E trigrams for a particular set were trigrams used as the C trigrams in other sets.

Separation of the three presentations of the acquisition-stage pairs was manipulated by inserting pairs and their cue trials from other sets, filler pairs and filler-pair cue trials, or test trials from other sets. With a separation interval of 1 (SI_1), the three presentations of each acquisition-stage pair were shown consecutively. With a separation interval of 3 (SI_3), the three presentations of the acquisition-stage pairs for a particular set were separated by the display of two pairs and cue trials which were from other sets or were fillers. With a separation interval of 6 (SI_6), five pairs and cue trials intervened between repetitions of the acquisition pairs for a particular set. The filler pairs were CVCs that were not used in constructing the pool of mediation and control sets.

The three separation intervals (1, 3, or 6) were combined factorially with the type of presentation (mediation or control) for each list. The resulting six cells were each represented six times among the 36 sets assigned to a particular list. Six lists were constructed so that over the entire experiment the 36 sets represented each of the six cells equally often.

All learning materials were stamped in capital letters on 4 × 6 in. index cards. For test-trial cards, Stimulus A was printed above the row of three response alternatives.

Experiment II used the same six lists of materials prepared for Exp. I. The only difference was that the second and third presentation of the acquisition pairs and their cue trials were reversed so that the cue trial preceded the pair. Similarly, the second and third repetitions of the second acquisition pairs were reversed. The first presentation of a pair was always followed by the cue trial.

Subjects.—The 24 introductory psychology students at Indiana University who participated in Exp. I were assigned randomly to the six lists with the restriction that all six lists be used before a list was repeated. The same restriction was imposed on the assignment of the 36 *Ss* from the same source who participated in Exp. II.

Procedure.—Three practice sets using different components in all positions were shown before the 36 experimental sets. On all sets, *Ss* read aloud the letters of the trigrams.

On cue trials, they were instructed to respond with the missing member of the pair. If *S* erred on the cue trial, the pair was presented again, then the cue trial was shown, etc., until *S* responded correctly. During the test trials, *S* successively paired each of the response alternatives with the stimulus and then selected the response which made the "best pair" with the stimulus.

RESULTS

Experiment I.—The six lists did not produce reliable differences in the mean numbers of selections of C on the test trials nor did the lists interact significantly with either the separation intervals or the type of presentation. The mean numbers of selection of C on test trials following both mediation and control presentation decreased as the separation intervals were increased, F (2, 36) = 14.92, $p < .01$, as shown in Fig. 1. In addition, more C alternatives were selected following mediation presentation than following control presentation (mediated facilitation) but the difference was short of statistical significance, F (1, 18) = 3.21, $p > .05$.

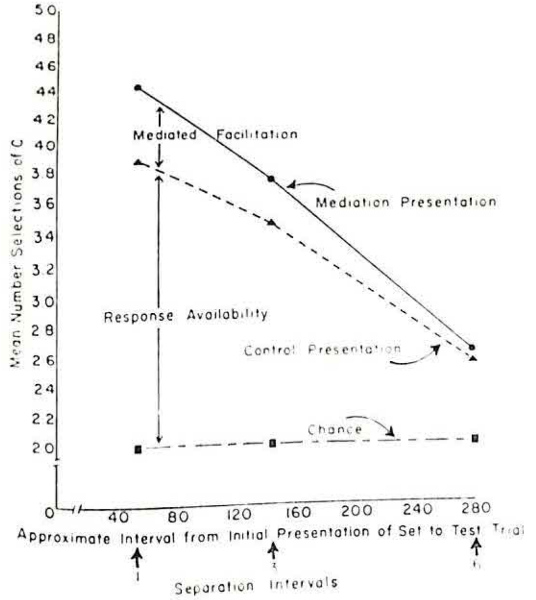


FIG. 1. Mean number of selections of C on the test trials: Exp. I.

As suggested by the profiles in Fig. 1, reliable evidence of mediated facilitation was found with massed presentation, t (36) = 1.98, $p < .05$, but not with either level of distributed practice. The t tests were computed using the residual error term with $df = 36$.

The linear trend of the profiles in Fig. 1 had a slope significantly different from zero, F (2, 36) = 15.58, $p < .01$, but the slopes of the linear trends for the two types of presentation did not differ appreciably, $F < 1$. The quadratic trends were not reliable.

Response availability was assessed by subtracting the number of C selections on the test trials expected if all three alternatives were equally likely from the number of selections of C actually made on test trials following control presentation. Reliable evidence of response availability was found at all levels of distribution of practice of the acquisition-stage pairs (the t for the separation interval with the smallest difference from the expected value, SI_6, was 1.93, $df = 23$). The reduction in the index of response availability as the distribution of practice of the acquisition-stage pairs increased was also reliable.

A grand total of 72 errors was made on the cue trials. Because of the infrequency of the errors in Exp. I no systematic consideration was given.

Experiment II.—The mean numbers of selections of C associated with the six lists did not differ significantly. The mean numbers of selections of C on test trials following mediation presentation with one, three, and six separation intervals and the corresponding means following control presentation are shown in Fig. 2. Increases in the distribution of practice of the acquisition stages were inversely related to the mean numbers

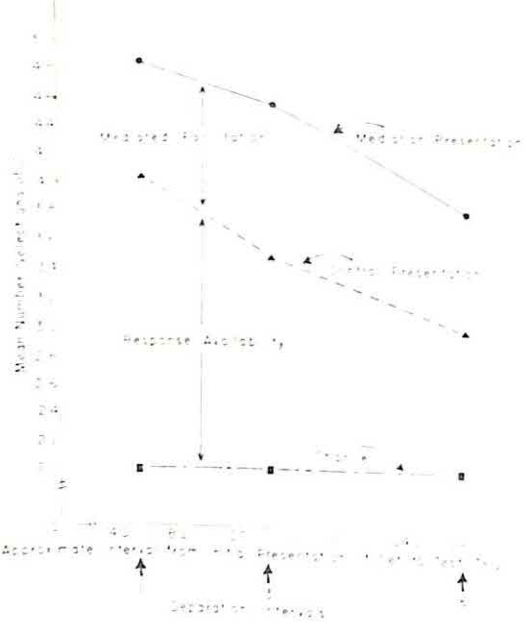


FIG. 2. Mean number of selections of C on the test trials: Exp. II.

of selections of C on the test trials, replicating the results of Exp. I, F (2, 60) = 12.82, $p < .01$. However, in Exp. II, mediated facilitation was found with all levels of distribution of practice, F (1, 30) = 14.40, $p < .01$, and was independent of distribution of practice. The t's for comparisons of the type of presentation were 9.48, 12.42, 9.48 for separation intervals of 1, 3, and 6, respectively. With $df = 60$, the p's $< .01$. These relationships are shown distinctly in Fig. 2. Tests for trend substantiated the above results: The linear slope of the two types of presentation combined differed from zero, F (2, 60) = 18.57, $p < .01$, but the linear trends of the two types of presentation evaluated separately must be considered essentially parallel, $F < 1$. Quadratic components were negligible.

Response availability decreased reliably as the separation intervals increased for all three separation intervals, similar to the Exp. I results. The t (35) for the difference between estimates of response availability with separation intervals of 1 and 6 was 3.23, $p < .01$.

The changed cue-trial technique employed in Exp. II produced many more errors than were found with Exp. I. The errors for Exp. II totaled 2,202, consisting of 1,891 omissions (no response given to the cue), 84 previous responses (responses which were appropriate to sets other than the one being tested), and 227 intrusions (the responses correct for the first acquisition-stage pair of the set tested, C, were substituted for the second acquisition-stage responses of the same set, A). The relative frequencies of the errors associated with type of presentation, separation interval, test-trial selection, acquisition stage, and blocks of 12 sets appear in Table 1.

The error data were expected to yield information about a number of relationships. The effects of type of presentation and separation interval were investigated first. Secondly, the errors associated with mediation-presentation sets whose test trials terminated with the selection of C were compared with the relative errors on mediation-presentation sets whose test trials ended with the choice of an alternative other than C. Presumably, mediation (plus response availability) occurred if C was selected following mediation presentation and did not occur if C was not selected following mediation presentation. The same analysis was performed on the control-presentation sets since response availability was assumed to operate if C was selected and to be inoperative if C was not selected. A third evaluation considered the possibilities of transfer effects both between and within sets.

The relative frequency of omissions

TABLE 1

Relative Frequencies of Omissions, Previous Responses, and Intrusions Made on Acquisition Pairs in Exp. II

SI	Mediation Presentation												Control Presentation											
	C Selected						C Not Selected						C Selected						C Not Selected					
	Blocks 1–12		Blocks 13–24		Blocks 25–36		Blocks 1–12		Blocks 13–24		Blocks 25–36		Blocks 1–12		Blocks 13–24		Blocks 25–36		Blocks 1–12		Blocks 13–24		Blocks 25–36	
	Acquisition Stage												Acquisition Stage											
	1st	2nd	1st	2nd	1st	2nd	1st	2nd	1st	2nd	1st	2nd	1st	2nd	1st	2nd	1st	2nd	1st	2nd	1st	2nd	1st	2nd
Omissions																								
1	.05	.12	.02	.05	.05	.05	.06	.06	.10	.10	.00	.06	.02	.14	.06	.02	.06	.07	.00	.10	.04	.08	.00	.00
3	1.39	.53	1.03	.60	1.07	.54	1.35	.70	1.29	1.00	1.19	1.00	1.41	1.05	1.08	.95	.98	.93	1.30	.05	1.03	.94	1.12	.92
6	1.20	.73	1.16	.64	1.11	.94	1.71	1.43	1.33	1.37	1.44	.96	1.49	1.24	.82	1.15	.87	1.27	1.46	1.06	1.12	1.31	1.29	1.43
Previous Responses																								
1	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.02	.02	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
3	.08	.02	.07	.07	.02	.00	.00	.04	.00	.00	.00	.06	.10	.00	.08	.05	.07	.00	.18	.00	.03	.00	.08	.00
6	.14	.02	.04	.04	.04	.02	.11	.00	.07	.00	.16	.04	.08	.00	.12	.05	.17	.00	.06	.00	.06	.00	.05	.00
Intrusions																								
1		.12		.00		.00		.06		.00		.00		.00		.00		.00		.00		.00		.00
3		.07		.06		.08		.07		.04		.03		.03		.08		.00		.00		.00		.04
6		.06		.05		.04		.11		.15		.24		.00		.02		.00		.00		.00		.00

on mediation-presentation sets was .70 and on control-presentation sets was .76, a nonsignificant difference. The values corresponding to separation intervals of 1, 3, and 6 were .06, .86, and 1.66, respectively. The sign tests conducted on these data because of the many zero frequencies for the single separation-interval conditions indicated that all of the differences were significant at the .001 level. The same reliable trend for omissions to increase with increased separation intervals was also observed when mediation and control sets were evaluated separately.

The second group of comparisons involved the omissions associated with the selection or nonselection of C on the test trials. Type of presentation interacted with test-trial behavior: mediation-presentation sets which terminated with the selection of C on their test trials yielded reliably fewer omissions, .68, than the mediation sets which ended with the choice of another alternative, .84 (sign test $p = .004$). Thus, fewer omissions were made on the mediation sets which showed facilitation. A Spearman rank-order correlation coefficient computed using the customary measure of mediated facilitation (the difference between critical responses on test trials following mediation and following control presentation) and the relative frequencies of omissions was $-.36$ (34), $.05 > p > .02$. However, the omission rate was not related to the test-trial behavior following control presentation sets—the relative frequencies of omissions for the control-presentation sets which did, or did not, terminate with the test-trial selection of C were .76 and .74, sign test $p > .05$ and ρ (34) between omissions and the measure of response availability was $-.17$, $p > .05$.

The third set of questions to be asked of the error data dealt with transfer effects. First, the possibilities of facilitory or inhibitory effects which extended over the experimental session were investigated by dividing the 36 sets into three blocks of 12 sets each. As Table 1 shows, a slight facilitory effect appeared to operate over the blocks of sets, i.e., omissions diminished over the experimental session except for the control-presentation sets whose test trials did not end with the selection of C. The second transfer effect, a within-set comparison, considered the likelihood that the frequencies of omissions on the first acquisition-stage cue trials might differ from the frequency of omission on the second acquisition-stage cue trials. And, indeed, fewer omissions were made on the second acquisition stage, particularly for the mediation-presentation sets whose test trials terminated with the selection of C (first stage = .79 and second stage = .47). Thus both the within-set and the between-set transfer effects were positive when omissions were scored.

The analyses of the previous responses showed some of the same trends: a tendency to increase with increases in the separation intervals and a tendency for more previous responses to be given during the first acquisition stage than the second acquisition stage. Initially the mediation-presentation sets which ended with the selection of C on the test trials produced more previous responses with separation intervals of 6 than with the other conditions but the frequency of previous responses dropped to about the same or slightly lower level than the other groups for later blocks of sets. The ρ's (34) for previous responses with mediated facilitation and response availability were $-.13$, $p > .05$; and .01, $p > .05$.

The third category of errors, the intrusions, was the most interesting because it was expected to yield information about proactive interference. If proactive interference operated, the response from the *first* acquisition pair would be given more frequently on the cue trials testing retention of the *second* acquisition pair with mediation presentation than with control presentation. The frequency of intrusions with mediation presentation sets, .07, reliably exceeded the intrusions for control presentation sets, .01, sign test $p = .006$.

Somewhat fewer intrusions occurred with mediation-presentation sets that ended with the test-trial choice of C, .05, than with mediation-presentation sets that did not, .08. The intrusions for mediation-presentation sets which terminated with the test-trial selection of C remained at about the same level for the three separation intervals ($SI_1 = .04$, $SI_3 = .07$, and $SI_6 = .04$) and decreased over the three blocks of sets (Sets 1–12 = .08; Sets 13–24 = .04; Sets 25–36 = .04). The intrusions for mediation-presentation sets which terminated with the choice of another alternative increased over the separation intervals ($SI_1 = .03$, $SI_3 = .06$, and $SI_6 = .11$) and remained at approximately the same level over the three blocks of sets (Sets 1–12 = .08, Sets 13–24 = .06, Sets 25–36 = .09). Intrusions occurred infrequently with the control-presentation sets.

DISCUSSION

Distribution of practice of the repetitions of acquisition pairs reduced response availability in both experiments. This result was consistent with previous work (Peterson et al., 1964) which showed that response availability was inversely related to the temporal interval existing between presentation of the C syllable during acquisition and its appearance on the test trial. In the present experiments this interval increased with the separation interval, thus reducing response availability.

The relationship between distribution of acquisition-pair practice and mediated facilitation was more complicated. The mean numbers of C alternatives chosen following mediation presentation decreased as distribution of practice of the acquisition pairs increased in both experiments. However, as depicted in Fig. 1 and 2, this measure was assumed to reflect the operation of two factors, mediated facilitation and response availability. When the contribution of response availability was removed by subtracting the mean numbers of selections of C following control presentation, mediated facilitation was inversely related to distribution of practice in Exp. I but in Exp. II, with an anticipation technique used to test retention of the second and third presentations of the acquisition pairs and with more *S*s per cell, mediated facilitation was unrelated to the distribution of acquisition-pair practice.

The differences between the estimates of mediated facilitation yielded by the two experiments might have been produced by a number of factors. The larger number of *S*s in Exp. II probably stabilized the observations somewhat, but this factor could not reasonably be invoked to explain the differences. Another possibility was that the slopes of the curves for the control conditions differed. Consequently, a *t* test was computed comparing the number of C selections made with a separation interval of 1 minus the C selections with a separation interval of 6 following control presentation for Exp. I with the same difference measure for Exp. II. The resulting value, $t\ (58) = .58$, $p > .05$, indicated that the overall slopes for the control conditions were not significantly different. A third candidate for explaining the diverse results of the two experiments was the retention-technique variation.

Some *E*s (e.g., Hawkins, 1964; Peterson & Brewer, 1963) have reported that procedures similar to prompting tended to facilitate learning more than anticipation procedures, particularly during early trials. While gross differences in learning materials and methodology rendered their conclusions largely inapplicable to the present case, extension of their results would predict lesser amounts of mediated facilitation with Exp. II rather than the greater amounts actually obtained.

Obviously, a more analytical comparison of the retention techniques was needed. When the temporal intervals elapsing between the presentations of a pair and its retention test were considered, striking differences emerged which were compounded by interactions with the separation intervals. Whenever the prompting method was used, the retention interval was approximately 2 sec., regardless of the separation interval assigned. Thus, the prompting method tested an immediate association, well within the customary limits of the immediate memory span, but did not necessarily provide evidence of a more lasting association. When *S* responded correctly on the cue trials of Exp. I, he demonstrated retention of the response for about 2 sec. and the experimental procedure was then resumed with the assumption that *S* had learned the association. The data suggested that *S* had learned the associations—to the extent that test-trial behavior was successfully mediated but only when the intervals between presentations of the components of a set were approximately 2 sec. These results were consonant with those of an earlier study (Peterson, 1965) which obtained mediated facilitation only when the intervals separating the first from the second acquisition pair and the second acquisition pair from the test trial did not exceed 2 sec. (with CVCs as the learning materials). However, in that study each acquisition pair was shown once.

In Exp. II, with the anticipation technique, the retention intervals were approximately 2, 18, and 45 sec. with separation intervals of 1, 3, and 6.[2] The longer separation intervals yielded information about the retention of the association for at least 18 or 45 sec.—and may have been testing somewhat different aspects of the retention process. Peterson (1963) suggested that some components may cue retention when intervals of up to 2 sec. are tested that are no longer present later in the retention interval. In Exp. II, we assumed that if *S* retained the pair for the tested interval during the acquisition stages he was more likely to retain the pair until the test trial. Support for this assumption was provided by Peterson, Saltzman, Hillner, and Land (1962) who showed that the longer the retention interval preceding a first test (up to 8-sec. intervals were tested) the greater the probability that the item would be retained on a second test. Thus it was not surprising that mediated facilitation was found at all levels of distributed practice when longer retention intervals were interspersed among the presentations of Exp. II.

In both experiments, pairs were reshown if *S* erred on the cue trial. One effect of the stipulation that all cue trials end with a correct response was that the more stringent retention tests of Exp. II resulted in more errors and, consequently, in more frequent exposures of the acquisition pairs than was the case for Exp. I. Thus more associations may have been learned in Exp. II through incremental strengthening of the associations during their more frequent exposures. Alternatively, more associations may have been learned because more opportunities for learning the pairs in an all-or-none fashion attended the additional exposures. The fact that mediated facilitation did *not* increase with the number of errors on the cue trials suggested discrete rather than incremental strengthening of the acquisition pairs.

When three categories of errors were

---

[2] On the average, about 9 sec. were required for *S* to respond to the presentation of a pair and its cue trial.

considered it became obvious that omissions and previous responses showed positive transfer from the cue trials testing the first acquisition pair, particularly for the mediation-presentation sets which showed facilitation by the selection of C on the test trials. Omissions increased as the separation intervals increased for all conditions.

The positive transfer reflected by the omissions and previous responses was not observed with the intrusions. Instead, intrusions were more frequent with mediation sets than with control sets, demonstrating the proactive interference predicted by Jenkins (1963). The mediation sets which evidenced facilitation by termination with the selection of C on the test trial were associated with fewer intrusions than the mediation-presentation sets which did not end with the selection of C. The question was Why?

The possibility that formal similarity among the nine letters comprising the three CVCs of a mediation set might have fostered proactive interference for some of the mediation sets and not for others was examined by comparing the relative frequency of selection of C on the test trials associated with the two sets for which there were two duplications ($\bar{X} = 13.50$), the nine sets with one duplication ($\bar{X} = 13.22$), and the 25 sets with zero duplications ($\bar{X} = 12.93$). These values negated the possibility that letter overlap was associated with reduced frequencies of the selection of C following mediation presentation.

Differences in the retention of the second acquisition-stage pairs associated with the mediation-presentation sets that demonstrated facilitation and those that did not were investigated by computing the relative frequencies of sets with no errors committed on the three presentations of the second acquisition-stage pairs. For the distributed-practice conditions a pair had to be retained for at least 18 or 45 sec. and be given correctly to the three cues for the set to be tallied. The relative frequencies of errorless second-stage acquisition sets for mediation-presentation sets followed by the selection of C on the test trials were .91, .23, and .20 for separation intervals of 1, 3, and 6. The corresponding frequencies for mediation-presentation sets which did not terminate with the selection of C were .74, .13, and .06. The second-stage acquisition pairs of the mediation sets demonstrating facilitation were retained more often than the second-stage acquisition pairs of the mediation sets not demonstrating facilitation for all separation intervals. The differences between the two groups of mediation sets were evaluated at each separation interval using sign tests, all p's $< .01$, supporting the earlier contention that pairs which were learned so that they were retained after periods of time were more likely to successfully mediate test-trial behavior than the associations which were not (at least demonstrably) retained.

Returning to Underwood and Schulz's (1961) statement that distribution of practice might facilitate learning when some interference existed in the response-learning phase, the data of the present experiments demonstrated that in spite of the transfer effects which were shown by the response members of the acquisition-stage pairs mediated facilitation was hindered by distributed practice of the acquisition-stage pairs in Exp. I, and was apparently independent of distributed practice in Exp. II. These results were not surprising since Underwood and Schulz did not find enhanced performance following distributed practice with trigrams or with words whose response members produced interference, but only with bigrams.

The most significant result of the present experiments was that distribution of the *retention tests*, combined with the stipulation that all retention tests end with a correct response, reduced response availability and demonstrated mediated facilitation of experimentally acquired associations over intervals totaling as much as 279 sec. from the initial presentation of the first acquisition pair to the test trial (SI_6).

REFERENCES

Goss, A. E. Verbal mediating responses and concept formation. *Psychol. Rev.*, 1961, **68**, 248–274.

Hawkins, J. R. The influence of training procedure and other task variables in paired-associate learning. *J. verbal Learn. verbal Behav.*, 1964, **3**, 70–76.

Jenkins, J. J. Mediated associations: Paradigms and situations. In C. N. Cofer & B. S. Musgrave (Eds.), *Verbal behavior and learning: Problems and processes.* New York: McGraw-Hill, 1963.

Peterson, L. R. Immediate memory: Data and theory. In C. N. Cofer & B. S. Musgrave (Eds.), *Verbal behavior and learning: Problems and processes.* New York: McGraw-Hill, 1963.

Peterson, L. R., & Brewer, C. L. Confirmation, correction, and contiguity. *J. verbal Learn. verbal Behav.*, 1963, **1**, 365–371.

Peterson, L. R., Saltzman, D., Hillner, K., & Land, V. Recency and frequency in paired-associate learning. *J. exp. Psychol.*, 1962, **63**, 396–403.

Peterson, M. J. Effects of delay intervals and meaningfulness on verbal mediating responses. *J. exp. Psychol.*, 1965, **69**, 60–66.

Peterson, M. J., Colavita, F. B., Sheahan, D. B., III, & Blattner, K. C. Verbal mediating chains and response availability as a function of the acquisition paradigm. *J. verbal Learn. verbal Behav.*, 1964, **3**, 11–18.

Underwood, B. J., & Schulz, R. W. *Meaningfulness and verbal learning.* New York: Lippincott, 1960.

Underwood, B. J., & Schulz, R. W. Studies of distributed practice: XXI. Effect of interference from language habits. *J. xp. Psychol.*, 1961, **62**, 571–575.

(Received July 7, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 6, 559-563

EFFECTS OF INTERTRIAL REINFORCEMENT ON RESISTANCE TO EXTINCTION FOLLOWING EXTENDED TRAINING[1]

ROGER W. BLACK AND KENNETH W. SPENCE

State University of Iowa

3 groups of 15 Ss each were trained to run a straight alley with food reward. 2 groups, partially reinforced experimental group (PR-X) and partially reinforced control group (PR-C), received 4 reinforced (R) trials, 4 nonreinforced (N) trials, and 2 intertrial reinforcements (ITR) daily for 12 days. PR-X received its ITRs after N trials preceding R trials. PR-C received ITRs following R trials. The continually reinforced group (CR) was trained with consistent reinforcement. PR-X and PR-C did not differ in extinction performance and both were more resistant to extinction than CR. The results are discussed in terms of the Hull-Sheffield hypothesis as revised by Capaldi and the frustration formulation of Amsel and Spence.

The recent experiments of Capaldi and associates (Capaldi, Hart, & Stanley, 1963; Capaldi & Spivey, 1963) have provided strong support for the Hull-Sheffield aftereffects hypothesis as to a possible factor responsible for the greater resistance to extinction of a learned response following partial reinforcement (PR) than after consistent reinforcement (CR). According to this formulation, the greater resistance to extinction of partially reinforced *S*s is that the stimulus aftereffect (S_n) of a nonreinforced trial (N) becomes strongly conditioned to the instrumental response, whereas under continuous reinforcement such conditioning does not occur as S_n is never present. Since extinction involves the aftereffect, S_n, on all trials beyond the first, the partially reinforced *S*s would have a greater excitatory strength for the instrumental response than the continuously reinforced *S*s and hence would require more trials to extinguish.

Modifying the Hull-Sheffield formulation, which assumed that the stimulus traces of reinforced (S_r) and nonreinforced trials (S_n) dissipate or undergo diminution as a function of time, Capaldi has assumed that a particular aftereffect will persist or remain functional indefinitely until replaced by another, different aftereffect. As a test of this modified conception Capaldi et al. (1963) suggested that the aftereffects of an N trial could be eliminated or replaced by giving *S* a subsequent direct placement in the baited goal box, a procedure described as "intertrial reinforcement" (ITR). An ITR following an N trial would thus replace S_n with S_r and, if an ITR is given following each N trial that precedes a reinforced (R) trial, S_n would never be reinforced and, hence, would not acquire a tendency to evoke the instrumental response. The *S*s trained under such PR conditions would not be expected to show greater resistance to extinction than *S*s trained with CR. In the Capaldi et al. (1963) study, ITR was given to two PR (50%) groups and a CR

---

[1] The writers express their appreciation to Patricia Elstad and Roy Matsumoto for their aid in the collection of these data.

(100%) group. One partial group was given ITR after R trials, the other after N trials. In agreement with theoretical prediction, the PR group receiving ITR following N trials did not differ significantly from the CR group. On the other hand, the PR group receiving ITR following R trials was significantly more resistant to extinction than either the CR group or the PR group that had S_n traces removed by ITR.

A similar finding has been reported by McCain (1964) in which only five trials involving but two N trials followed by R trials were given in the acquisition period. As the result of his findings, McCain raised certain questions with regard to the frustration-type theory of the PR effect offered by Amsel (1958) and Spence (1960). This latter theory makes use of the concept of a conditioned frustration response (r_f) and the conditioning of the cues (S_f) resulting from this response to the instrumental response. Since this theoretical mechanism is, in turn, assumed to be dependent upon the prior conditioning of an anticipatory goal response (r_g) to the stimulus situation (e.g., runway), it is evident that this theory would not expect greater resistance to extinction with such a brief sequence of PR trials as McCain employed.

Examination of the empirical evidence as to the number of acquisition trials required to produce the PR effect indicates that the findings are highly dependent upon the magnitude of the reward employed. With relatively small rewards investigators have failed to obtain the effect with 16 acquisition trials (Wagner, 1961), 20 (Yamaguchi & Sukenmune, 1963), 24 (Amsel, 1958), and 30 (Hulse, 1958; Weinstock, 1957). On the other hand, when relatively large rewards were used, enhanced resistance to extinction under PR has been obtained with only 16 acquisition trials (Wagner, 1961), and McCain has reported a PR effect after as few as 3 (McCain, Love, & Greur, 1962) and 5 acquisition trials (McCain, 1964). These latter findings are somewhat surprising inasmuch as only one and two reinforced S_n R sequences were involved in the respective studies.

While the dependency of the PR effect on the magnitude of reward can be nicely handled by the frustration theory, its occurrence after as few as 3–5 acquisition trials is definitely not to be expected.

Confirmation of these McCain studies would require the assumption of some other factor than the frustration mechanism. One possibility is that both hypothetical factors, stimulus aftereffects and frustration, play a role in producing the PR effect. Thus, the aftereffects factor might be conceived as developing first and the frustration factor after more prolonged training.

The frustration interpretation implies that if a sufficiently large number of acquisition trials are given to insure that frustration occurs on N trials and becomes conditioned as specified by the theory, then *Ss* trained under PR with ITR after either N or R trials should exhibit greater extinction than *Ss* trained with 100% reinforcement. The present study tests this implication of the frustration theory by extending the acquisition period to 96 trials, a period of acquisition that has been shown to produce the PR effect even when very small rewards have been employed.

METHOD

Subjects.—The *Ss* were 45 male, hooded rats from the colony maintained by the

Department of Psychology of the State University of Iowa and were 90–100 days old at the beginning of training.

Apparatus. The apparatus was a straight alley constructed of wood and consisting of a start box 10 in., a runway 3 ft., and a goal box 8 in. in length. All segments of the alley were $3\frac{3}{4}$ in. wide and high and were painted black. The start box was separated from the runway by a guillotine start door and the goal box could be separated from the runway by a retrace door. The alley was covered by a clear Plexiglas, hinged top.

Opening the start door started a Hunter klockounter which operated until *S* interrupted a light beam to a photocell located 6 in. past the start door. Interruption of this light stopped the first klockounter and started a second which operated until *S* interrupted a light located 6 in. farther along the alley. Interruption of a third light located 6 in. before the goal box started a final klockounter which stopped when *S* reached a light located 6 in. within the goal box. Thus, starting, running, and goal-box times, respectively, were recorded.

Pretraining.—Ten days prior to the beginning of training all *Ss* were placed on a maintenance schedule consisting of 12 gm. of standard lab chow given each day at about the time their subsequent daily training trials would be finished. Water was continuously available. On each of the final 3 days prior to the beginning of training, *Ss* were placed in individual carrying cages, taken to the experimental room, and allowed to explore the unbaited alley individually for about 5 min.

Experimental procedure.—The *Ss* were randomly divided into three groups of 15 each, designated partially reinforced experimental group (PR-X), partially reinforced control group (PR-C), and continually reinforced group (CR). Eight acquisition trials were given daily for 12 days. For PR-X and PR-C, half of the acquisition trials were reinforced (R) trials and half were nonreinforced (N) trials. On an R trial, *S* was placed in the start box and, when *S* oriented to the start-box door, that door was opened. A caster filled with wet mash was present in the goal box and when *S* entered the goal box, the retrace door was closed and *S* was allowed access to the wet mash for 20 sec. An N trial was conducted in an identical fashion except that the food dish was empty. For the CR group, all trials were reinforced. The *Ss* were run sequentially in squads of nine which provided an intertrial interval (ITI) of about 10 min.

The sequence of trials each day was RNNRRNNR. This provided two cases daily in which an N trial preceded an R trial—i.e., Trials 3 and 7. Each *S* also received two intertrial reinforcements (ITR) daily. For the PR-X group, the ITRs were given following N trials which preceded R trials—i.e., Trials 3 and 7. For PR-C and CR, the ITRs followed R trials—specifically, Trials 4 and 8. An ITR consisted of replacing *S* in the baited goal box for 20 sec., 15 sec. after it had been removed from the goal box at the end of the previous trial.[2] The *S* was then returned to its carrying cage to await the next trial.

Extinction for all groups consisted of 21 consecutive N trials given on the thirteenth day of training, at an ITI of 8–12 min. The procedure was identical to that employed during acquisition with the exception that no ITRs were given.

RESULTS AND DISCUSSION

Individual starting, running, and goal-box times were converted to speeds (ft/sec). Since analysis indicated that the results for each measure were essentially the same, only the running-speed data are discussed. Figure 1 presents the extinction curves in blocks of three trials for the three groups. It is clearly evident that the two PR groups did not differ and that both were more resistant to extinction than CR. A Trial × Groups analysis of variance revealed the differences between the groups to be reliable, $F\ (2, 42) = 8.82$,

---

[2] The study was run in two replications with approximately the same number of *Ss* in each group for each replication. In the first replication, *Ss* were confined in the unbaited goal box for 20 sec. on N trials. Following a personal communication from Capaldi to the effect that 20 sec. of nonreward may produce an aftereffect too potent to be replaced by an ITR, the nonreward confinement time was reduced to 10 sec. Statistical analyses of the two replications failed to indicate any significant differences, however; and hence, the replication factor has been ignored.

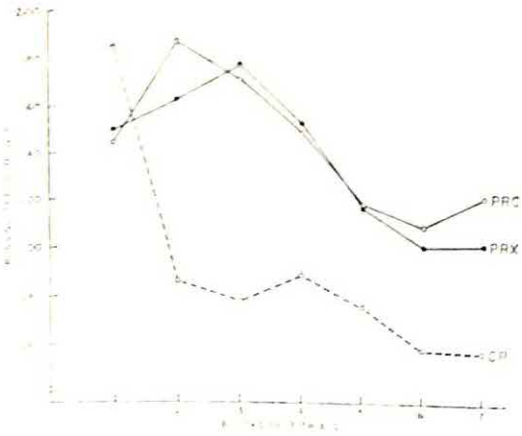


FIG. 1. Running speed during extinction as a function of blocks of three trials.

$p < .01$. The Groups × Trials interaction, evident in the crossover of the PR and CR curves, was also significant, $F\ (40,\ 840) = 4.12$, $p < .001$, but a comparison of PR-C and PR-X indicated no reliable difference, $t\ (43) = -.42$.

The results of the present experiment are thus seen to differ from those of Capaldi et al. (1963) and McCain (1964). Unlike the findings of these latter studies, the group (PR-X) that had ITRs after N trials did not, like the continuously reinforced group (CR), extinguish quickly, but rather, was like the partially reinforced control group (PR-C) in being highly resistant to extinction. These findings are in accord with the frustration theory in that both PR groups exhibited the expected greater resistance to extinction.

Considered together, the results of Capaldi et al., McCain, and the present experiment suggest that when the number of acquisition trials is small, the opportunity for the instrumental response to become conditioned to the aftereffects of nonreinforcement may be a determinant of resistance to extinction. However, when acquisition training is sufficiently extended to insure, theoretically, the occurrence of frustration on nonreinforced trials, extinction performance is in accord with the implications of the frustration hypothesis. This suggests that, once conditioned frustration is developed, its influence on performance is sufficiently great that the role of differential stimulus aftereffects becomes negligible.

Neglected as yet in these ITR studies is the variable, magnitude of reward. All of the studies employing this technique that have been reported thus far have involved relatively large rewards. As we know, when small rewards are employed partial reinforcement does not lead to greater resistance to extinction even up to as many as 30 acquisition trials. Such a finding implies that differential stimulus aftereffects are not operative early in acquisition with such small reward magnitude as they appear to be with large rewards.

REFERENCES

AMSEL, A. The role of frustrative nonreward in noncontinuous reward situations. *Psychol. Bull.*, 1958, 55, 102–119.

CAPALDI, E. J., HART, D., & STANLEY, L. R. Effect of intertrial reinforcement on the aftereffect of nonreinforcement and resistance to extinction. *J. exp. Psychol.*, 1963, 65, 70–74.

CAPALDI, E. J., & SPIVEY, J. E. Effects of goal-box similarity on the aftereffect of nonreinforcement and resistance to extinction. *J. exp. Psychol.*, 1963, 66, 461–465.

HULSE, S. H. Amount and percentage reinforcement and duration of goal confinement in conditioning and extinction. *J. exp. Psychol.*, 1958, 56, 48–57.

MCCAIN, G. The effect of intertrial reinforcement on resistance to extinction. Paper read at Midwestern Psychological Association, St. Louis, April 1964.

MCCAIN, G., LOVE, N., & GREUR, W. Extinction as a function of a small number of partially reinforced trials. *Psychol. Rep.*, 1962, 11, 451–454.

SPENCE, K. W. *Behavior theory and learning.* Englewood Cliffs, N. J.: Prentice-Hall, 1960.

WAGNER, A. R. Effects of amount and percentage reinforcement and number of acquisition trials on conditioning and extinction. *J. exp. Psychol.*, 1961, 62, 234–242.

WEINSTOCK, M. B. A factorial study of some variables affecting resistance to extinction under partial reinforcement with spaced trials. Unpublished doctoral dissertation, Indiana University, 1957.

YAMAGUCHI, M., & SUKENMUNE, S. Effects of total amount of reinforcement under partial reinforcement schedules on running speeds. *Jap. psychol. Res.*, 1963, 5, 137–141.

(Received July 10, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 6, 564-569

PARTIAL VISUAL FEEDBACK OF COMPONENT MOTIONS AS A FUNCTION OF DIFFICULTY OF MOTOR CONTROL

JOHN D. GOULD AND AMY SCHAFFER

IBM Watson Research Center, Yorktown Heights, New York

This research studied the interrelations of visual perception and movement. Closed-circuit television techniques allowed the joint action of S's hand, control instrument, and operational effects to be visually fed back singly or in various combinations. 2 levels of difficulty of each of the 3 task components responsible for the 3 types of visual feedback were varied independently of feedback in a task where S steered a ball through a maze. Results showed vision of the tool to be most important followed by vision of operational effects and hand-arm movements. No significant difference was found on the task-difficulty variable which may have influenced the absence of a significant interaction between type of visual feedback and movement difficulty.

It has been proposed (Gould, 1965a, 1965b; Smith, 1962) that the sensory feedback which occurs in all modalities in a complex motor-control task may be considered under three headings: that sensory feedback resulting from movement of the body, torso, arm, foot, or hand (called reactive feedback); that feedback resulting from movement of the tool, control, or instrument (called instrumental feedback); and that feedback resulting from action of the control or tool on objects or material in the environment (operational feedback). These three feedback effects may be observed in all sensory modalities, and which modalities are actually involved depends upon the overall motion pattern and the nature of the task itself.

In a previous study (Gould, 1965a) in which *S*s used a pliers to move pins into appropriate holes and viewed their performances on the monitor of a closed-circuit television system, a technique was developed whereby *S* could see his hand, tool, and environmental changes either singly or in various combinations. The nonadditive effect of reactive, instrumental, and operational *visual* feedback resulting from these three task components, respectively, was demonstrated. While the results of that study were interpreted to support a hypothesis that the relative effects of each of the three types of visual feedback are dependent upon the relative degree of complexity or difficulty of the motion patterns involved, this interpretation was not above criticism. That is, while visual feedback was varied, only one level of difficulty or complexity of the effect responsible for each type of feedback was employed. In the present study, however, the level of difficulty of the three components of the task itself was varied independently of variations in the amount or type of visual feedback. Accordingly, *S*s used two types of arm-hand movements in manipulating two different tools to control and steer a rolling ball (two speeds) into an appropriate slot. Performance at each of these eight task conditions, then, was evaluated for seven different visual-feedback conditions.

METHOD

The experimental task consisted of using a rod to steer a $\frac{1}{2}$-in. ball, that rolled down a 17-in. inclined alley, into the appropriate one

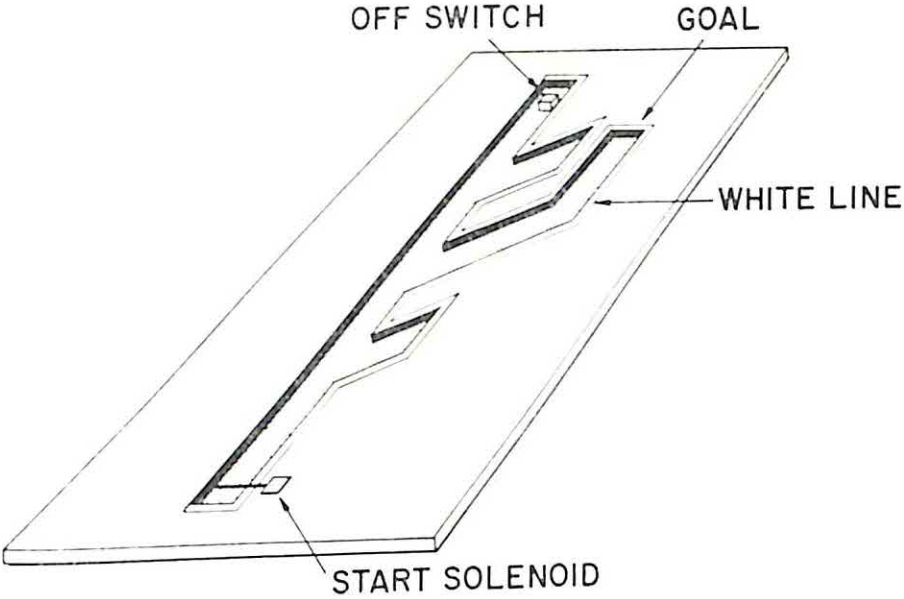


FIG. 1. Experimental maze.

of three $\frac{3}{4}$-in. maze arms, each of which intersected with the main alley as shown in Fig. 1. The alleys, shown as dotted areas, were $\frac{3}{4}$ in. wide and $\frac{1}{4}$ in. deep. The entire alley board was inclined so that the ball would roll down toward the off switch.

As will be described, the maze was located on the right side of *S*, since he viewed his performance on a television monitor located directly in front of him. Consequently, in one level of hand-arm movement, *S* used his right (preferred) hand; whereas in the other level, *S* used his left (nonpreferred) hand and reached across his chest. Two types of tools or instruments were used to steer the ball. In both cases, a 4-in. rod, $\frac{1}{2}$-in. diam., was mounted normal to an 8-in. handle. However, the end of one tool was convex while the end of the other tool was concave, the latter presumably allowing *S* better control over the rolling ball. The speed with which the ball would roll freely down the alley when released was varied to change the level of difficulty of that task component yielding operational feedback. This was accomplished by tilting the alley to a slope of .16 or .24 with respect to the table that supported it. When the ball was allowed to roll freely down the 17-in. main alley (during the experiment, of course, it was not) it took .91±5% sec. and 1.12±5% sec., respectively. This particular task was selected as one that would minimize any interaction between the three movement variables which would lead to a drastic altering of the general task.

To provide visual feedback of only one or combinations of the component movements, a high-resolution, closed-circuit television system was used. The *Ss* did not look directly at the maze, but instead viewed their performance on a television monitor (Conrac Model CN A8/C). A television camera (General Electric Model 4TE15B2) with a 1-in. lens was mounted in front of and perpendicular to the work board. By turning down the brightness and beam controls of the television system and elevating contrast and target abnormally the picture on the monitor was such that only white objects could be seen by *S* while black stimuli were not differentiated from the background. Accordingly, depending upon the feedback condition, *Ss* wore a black or white glove and used a black or white tool to control a black or white ball.

Seven visual-feedback conditions, as defined by what *S* saw, were used: vision of hand, tool, and ball or, in the terminology of this paper, visual feedback of reactive-instrumental-operational effects (RIO), vision of hand and ball only or reactive-instrumental (RI); instrumental-operational (IO); reactive-operational (RO); reactive only (R); instrumental only (I); and operational only (O).

A thin white stripe was painted adjacent to the upper edges of each alley, as shown in Fig. 1, and two small white spots indicated, respectively, the position of the "off" microswitch which *S* pushed with his tool upon completion of a trial and the goal box of the maze. Accordingly, the picture on the television screen that *S* saw was entirely black except for the white stripe surrounding the maze alleys and the white indicating spots, plus the additional feedback provided in a particular experimental condition. (For example, in the IO feedback condition, in addition to the stripes and indicating spots, the tool was seen as controlling the visible ball, although *S*'s hand was invisible.)

At the outset of each trial, the ball was positioned at the top of the main alley and held by a solenoid. The *S* positioned his tool on the floor of the alley 1 in. in front of the position where the ball was released. At a given signal, *E* threw a switch which released the ball and automatically started a time clock, accurate to .01 sec. The *S* then trapped the ball with his tool and steered it through the maze; upon completion he stopped the clock by hitting the microswitch with his tool. Correction for errors was not allowed, and *S*s were strongly instructed to minimize errors, as their time scores would be weighted for such errors. Both time scores and errors were recorded. Even on trials in which *S* made an error he still hit the microswitch which stopped the timer.

Eight *S*s, paid volunteers from a nearby college, were used in this experiment. Each *S* performed five trials in each of the eight movement conditions and each of the seven feedback conditions. An 8×8 Latin square formed the main statistical design, with *S*s as the row variable and experimental sessions as the column variable; movement conditions were varied across sessions. Within each of the 64 cells of the Latin square, the seven feedback conditions were randomized, and five consecutive trials were recorded at each feedback condition. Thus, six orthogonal variables—hand-arm movement (2 levels), type of tool (2 levels), ball speed (2 levels), visual feedback (7 levels), trials (5), and *S*s (8)—were represented in the analysis-of-variance design.

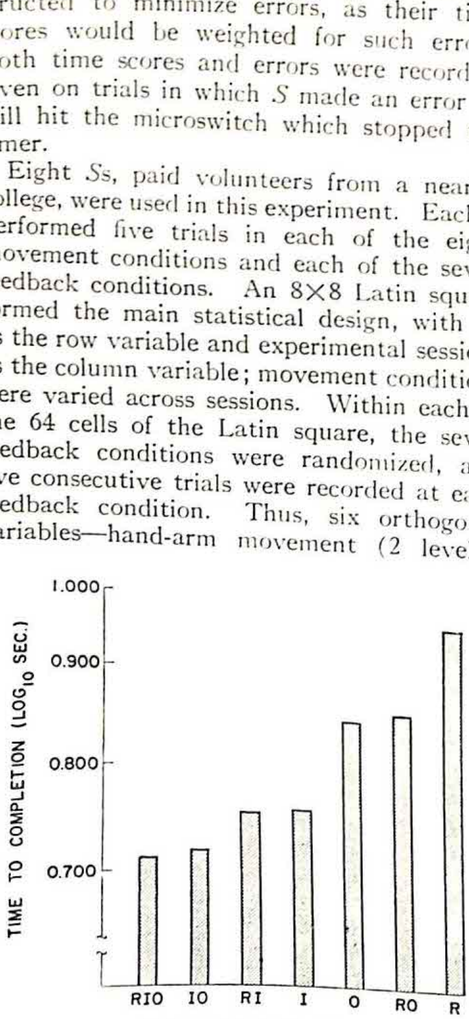


FIG. 2. Differences in performance as a function of visual feedback.

RESULTS

Both the 2240 time scores, converted to log values, and also the error scores were completely analyzed. The direction and statistical significance of the results were generally the same. However, occasional "wild" scores were involved, and, accordingly, to obtain better estimates of experimental means the median time score among the five trials for each *S* at each of the 56 experimental conditions was selected and analyzed. The results, which did not differ in terms of significance or direction of means from the unselected analysis, are presented here. It should be mentioned also that several weighting schemes were attempted in adjusting the time scores for errors but these did not appreciably change the direction or significance of any main effects or interactions.

Effects of visual feedback.—By varying the components of movement which were fed back, performance was differentially affected, $F\ (6, 42) = 64.33$, $p < .001$. In general, vision of the instrument was most helpful to performance followed by vision of the ball and vision of the hand. Figure 2, where average time to completion in $\log_{10}$ sec. is plotted on the ordinate with feedback conditions arranged in order of magnitude on the abscissa, shows that vision of the instrument was the most important in this task. That is, those four conditions in which instrumental feedback was provided (RIO, IO, I, and RI) were the easiest four. Operational feedback or vision of the ball was next most important, followed by vision of the hand or reactive feedback.

In originally applying range tests to the differences between the seven feedback means, it was found that the R feedback condition differed significantly from the other six, which did not differ significantly among themselves. However, the scores around the R mean were significantly more variable than were the scores around any of the other six feedback means, *F* (63, 63) = 1.85, for the smallest difference; $p < .02$. Accordingly, another range test was applied eliminating the R feedback condition and its associated error variance from the estimate of the overall MS_e. In this analysis, the means fell into the following two groups (RIO, IO, I, and RI) and (I, RI, O, and RO). Means which are inside the same parentheses did not differ from each other. Means not included within a particular group differed from that group at the .05 level of confidence.

Three results are evident from this analysis. First, the additional visual feedback of *S*'s hand did not affect performance significantly. Second, the conditions with feedback of the moving ball (RO and O) differed significantly from those conditions where *S* saw both the ball and his instrument (RIO and IO). Third, the effects of visual feedback of the ball and instrument separately were additive when combined, a fact that was not true in a previous study (Gould, 1965a).

Effects of movement conditions.—Overall, the eight movement conditions did not differ significantly from each other ($F < 1.00$). This was true when the effects of order of presentation were removed also. By breaking down the movement conditions into single degree-of-freedom comparisons, it was found that the two levels of ball speed contributed the most variance to the movement conditions, *F* (1, 7) = 3.94, $p < .10$. When these eight movement conditions were rank ordered, those four in which the ball rolled faster were all better than those four in which the ball rolled slower. No significant interaction occurred between the movement conditions themselves, a result which was essential in demonstrating that the basic nature of the task did not change much when each component was varied.

Movement × feedback interaction.—When the means for the Movement Conditions × Feedback interaction were plotted, the results showed that the effects of various amounts of visual feedback were nearly the same regardless of the movement or task condition. Statistically the interaction was not significant, *F* (42, 294) = 1.23. Since the only movement condition that approached significance was the effect of ball speed, the two curves for ball speed were plotted as a function of feedback. In the four cases where the ball or operational effects were not seen, the differences between the curves were larger than where the ball could be seen, although the interaction of Ball Speed × Feedback was not significant. Thus, not only was there no type of overall interaction between the task variables and visual feedback, but the presence or absence of vision of a particular task component did not interact with the difficulty of that component either.

DISCUSSION AND SUMMARY

The main results of this study, using a television system to control visual feedback of component movements, indicated that performance was dependent upon visual feedback, but that the levels of the task variables selected did not affect performance nor did they interact with visual feedback. Observation indicated that not only performance but the action and attitude of *Ss* differed greatly depending upon the feedback condition. In the four conditions providing instrumental feedback *Ss* generally moved their tool down the runway allowing the ball to roll after it, stopped at the second alley, and then steered the ball into that alley. Errors (8–11%) were relatively few; in fact, time scores and mean errors on the seven feedback conditions correlated highly (tau = +.90). Without instrumental feedback, maneuvering the ball into the alley became difficult, *Ss* sometimes took longer to find and push the termination switch, and errors doubled. The task became nearly impossible when *Ss* saw only their hand and arm (R); the greatest difficulty appeared to be in locating the appropriate arm of the maze. Errors occurred on 60% of the trials, and time scores would have been even longer except that once *S* perceived that an error was committed he did not correct it. Postexperimental observations indicated that R feedback and a completely blind condition were of the same degree of impossibility. In all conditions *Ss* only rarely experimented with new, invariably unsuccessful techniques, such as letting the ball roll freely until it hit the tool, stationed at the alley.

The results concerning visual feedback, where it was shown that vision of the tool was the most helpful to performance, followed by vision of the ball and then vision of the hand, are interesting in light of a previous study (Gould, 1965a) on a different task showing the order of importance of visual feedback to be operational, instrumental, reactive. In addition, a third study (Smith & Smith, 1962) using different techniques and a still different task studied three feedback conditions (RIO, RI, and O) and showed no difference between them. These results, based upon this experiment which emphasizes accurate timing of movements, as well as the other two experiments which emphasize spatial factors, indicate that the need for various components of visual feedback is dependent upon the general nature of the task or motor patterns involved. One implication of this conclusion concerns knowledge of result feedback (KR), traditionally considered as having reinforcement value and/or essential for motor learning, and generally considered synonymous only with what is here called operational feedback. Experiments, e.g., on differences between pursuit and compensatory tracking displays (Adams, 1961; Chernikoff & Taylor, 1957) or precision of displayed feedback (Hunt, 1961), vary operational KR and these studies show that the relative value of operational feedback depends upon the motor task involved; the present study demonstrates that in some tasks visual instrumental feedback is even more important than visual operational feedback.

Other experiments where feedback is impoverished or taken away (Abbey & Cowan, 1960; Diehl & Seibel, 1962; Hunt, 1961; Smith, 1963), studies of augmented feedback (Gould & Smith, 1964; Kinkade, 1963), and closed-circuit television studies of motor tasks performed in a variety of displaced and distorted visual-feedback conditions (summarized in Smith & Smith, 1962) all point to the general conclusion that the effects of type and amount of visual feedback on level of performance, rate of learning, and adaptation to visual distortion depend upon the motion pattern and, in the case of adaptation to one type of prismatically displaced vision (Held & Freedman, 1963), the movements of *S* during the experiment. In general, in relatively simple tasks where the degrees of freedom of movement patterns yielding multimodal operational, instrumental, and reactive feedback are relatively limited (Ryan & Bilodeau, 1962; Smith, 1963) or when the motor skill is extensively practiced, as in the case of a one-dimensional tracking situation (Hunt, 1961), or is well learned, as in typing (Diehl & Seibel, 1962), the effects of varying visual feedback are not as disastrous as in studies of complex motor patterns, as in the present experiment and others

(Abbey & Cowan, 1960; Gould, 1965a). In striving to develop a more precise understanding of the complex relationships between sensory feedback in all modalities and human or animal motion, the television technique used here offers much potential in providing partial visual feedback, and has application in experiments on differential adaptation of various component motions, parts of the body, or visual field in distorted or displaced vision, in the training of complex motor skills, and in studies where the intent is to provide *S* with limited body or environmental perception (e.g., Bossom, 1964).

REFERENCES

Abbey, D. S., & Cowan, P. H. Incomplete visual feedback and performance on the Toronto Complex Coordinator. *Percept. mot. Skills*, 1960, **11**, 43–45.

Adams, J. Human tracking performance. *Psychol. Bull.*, 1961, **58**, 55–79.

Bossom, J. Mechanism of prism adaptation in normal monkeys. *Psychon. Sci.*, 1964, **1**, 377–378.

Chernikoff, R., & Taylor, F. V. Effects of course frequency and aided time constant on pursuit and compensatory tracking. *J. exp. Psychol.*, 1957, **53**, 285–292.

Diehl, M. J., & Seibel, R. The relative importance of visual and auditory feedback in speed typewriting. *J. appl. Psychol.*, 1962, **46**, 365–369.

Gould, J. D. Differential visual feedback of component motions. *J. exp. Psychol.*, 1965, **69**, 263–268. (a)

Gould, J. D. Stereoscopic television tracking. *J. appl. Psychol.*, 1965, **48**, 369–377. (b)

Gould, J. D., & Smith, K. U. Sensory feedback analysis of stereo-television pursuit tracking. *J. appl. Psychol.*, 1964, **48**, 152–160.

Held, R., & Freedman, S. J. Plasticity in human sensorimotor control. *Science*, 1963, **142**, 455–462.

Hunt, D. P. The effect of the precision of information feedback on human tracking performance. *Hum. Factors*, 1961, **3**, 77–85.

Kinkade, R. G. A differential influence of augmented feedback on learning and performance. *USAF AMRL tech. docum. Rep.*, 1963, No. 63-12.

Ryan, F. J., & Bilodeau, E. A. Countertraining of a simple skill with immediate and 1-week delays of informative feedback. *J. exp. Psychol.*, 1962, **63**, 19–22.

Smith, A. H. Effects of continuous and intermittent feedback on precision in applying pressure. *Percept. mot. Skills*, 1963, **17**, 883–889.

Smith, K. U. *Behavior organization and work.* Madison: Univer. Wisconsin Press, 1962.

Smith, K. U., & Smith, W. M. *Perception and motion.* Philadelphia: W. B. Saunders, 1962.

(Received July 11, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 6, 570-574

GRAMMATICAL AGREEMENT AND SET IN LEARNING AT TWO AGE LEVELS[1]

JAMES G. MARTIN, JUDY R. DAVIDSON, AND MYRNA L. WILLIAMS

Chico State College

In Experiment I nouns were linked with verbs in paired-associate learning. Nouns were preceded by an article to induce a "sentence set." Grammatical number agreement was varied by inflecting either the noun or verb. ½ the Ss, 3rd, 5th, and 6th graders, and college students, learned a grammatical list, ½ an ungrammatical list. Grammatical lists were learned faster than ungrammatical lists by college Ss. In Experiment II ½ the noun-verb pairs were preceded by an article to induce a sentence set while ½ were not. ½ the pairs agreed in number. The pairs, making 4 conditions, were presented on sound tape to 1st and 2nd graders and college Ss. Grammatical pairs were learned faster at both age levels. Sentence set had no effect on grammatical pairs, but when combined with ungrammatical pairs the sentence set interfered with learning by college Ss but not by grade school Ss.

Writers on language behavior, whether challenging associationistic interpretations (e.g., Chomsky, 1957; Lashley, 1951; Miller, Galanter, & Pribram, 1960) or advancing them (e.g., Osgood, 1957), have emphasized the importance of integrative processes and determining tendencies or sets in the grammatical selection of words used or expected in a sentence. Thus the expectation of, say, a plural verb in a speech utterance is determined by an inflection indicating plural number in the noun which precedes it. Through such experience with the language, encoding and decoding habits and sets are developed which mirror those of the speech community.

This study is concerned with some implications of these notions of grammatical behavior in a conventional learning task. First, nouns or noun phrases were linked with verbs by the paired-associates method. If the presence of noun inflections calls grammatical habits into play, noun-verb pairs which agree in number (making a grammatical sentence) should be learned faster than ungrammatical pairs. Secondly, the stronger the grammatical set or expectation the greater should be the facilitation if materials are grammatical and the greater the interference if they are not. Thirdly, since grammatical habits and set develop through experience they should make more difference for older than younger *S*s.

In Exp. I, college and grade school *S*s learned noun-verb pairs which differed in number agreement. In Exp. II, in addition to varying grammatical agreement the pairs were constructed in such a way as to vary "sentence set." The appearance of the word THE preceding the noun should facilitate the learning of the pair if grammatical, or interfere with the learning of the pair if ungrammatical, or both. In Exp. II *S*s were from two age levels also. The results of the experiments indicate a resemblance between the effects of gram-

---

[1] Supported in part by Grant MH 08661-01, National Institute of Mental Health, United States Public Health Service, to the first author. Appreciation is due to Ted DeBernardi, Principal, and the teachers of the Amer Jay Hamilton School for their cooperation in this research.

matical variation and the more conventional variables of the rote-learning laboratory.

EXPERIMENT I

Method

Subjects.—There were 76 college *Ss* and 76 grade school *Ss*. The grade school *Ss*, 24 each from the fifth and sixth grades and 28 from the third grade, were from the college laboratory school. One failed to learn the practice list within 30 min. and was replaced by another *S*. The college *Ss* were from introductory psychology classes and participated for extra credit or as a class requirement. All *Ss* spoke English as their native language and some had been in verbal-learning experiments before. The *Ss* within each grade or age level were randomly assigned to lists in blocks of four.

Materials.—Five basic three-word sentences each consisting of a noun phrase and a verb were constructed. The noun phrases served as stimulus items in paired-associate learning, the verbs as response members. All were Thorndike-Lorge (1944) A or AA words. The semantic relations between nouns and verbs were chosen to be unlikely. Table 1 gives the sentences.

TABLE 1

BASIC SENTENCES IN EXP. I

THE EGG-HOPES
THE RAIN-JUMPS
THE SHIP-TALKS
THE BANK-PULLS
THE ROAD-GROWS

Each sentence was altered and arranged to yield four pairings, two grammatical and two not; they are most easily described by example: THE EGG-HOPES, THE EGGS-HOPE, THE EGG-HOPE, THE EGGS-HOPES. Four lists were constructed each with the five sentences in one of the altered forms making two grammatical and two ungrammatical lists. On each list the five pairs making a trial were randomized five times. Order was exactly parallel across the four lists.

Apparatus and procedure.—Lists were presented on a memory drum set behind a black screen. A light placed on the drum illuminated the aperture. The room was dimly lit.

The college *Ss* were given standard anticipation instructions for paired associates. Instructions to the grade school *Ss* were modified. Prior to the instructions for learning each grade school *S* was given a card on which appeared each word (from all four lists) and was asked to read the words to a criterion of one correct reading in three trials (with help from *E* on the first two trials). All *Ss* met the reading criterion.

Exact reproduction of the responses was emphasized in the instructions for all *Ss* and was scored accordingly. Pairs were presented at a 2:2-sec. rate with a 4-sec. interpair interval (IPI) and a 16-sec. intertrial interval (ITI). A practice list of two nonsense pairs, THE GUZ-WIK and THE REG-DAT, was learned to one perfect trial. Then the test list was presented till the criterion of one correct trial was reached or for a maximum of 30 trials. Data from *Ss* who reached maximum trials were included. One fifth grader and three third graders reached 30 trials on a grammatical list, four third graders reached the maximum on an ungrammatical list, and thus the means for their corresponding groups are underestimated.

Results

Table 2 shows the trials to criterion for each age level and type of list.

TABLE 2

TRIALS TO CRITERION IN EXP. I

List	Grade Level								*M*
	Third		Fifth		Sixth		College		
	N	*M*	*N*	*M*	*N*	*M*	*N*	*M*	
Grammatical	14	15.21	12	8.92	12	6.59	38	6.08	8.29
Ungrammatical	14	17.71	12	8.83	12	9.00	38	9.13	10.64
M	28	16.46	24	8.87	24	7.79	76	7.62	

TABLE 3

BASIC SENTENCES IN EXP. II

BANKS-PULL
STONES-WIN
COATS-STOP
SHIPS-TALK
SHOES-EAT
DAYS-CUT
BOOKS-LAUGH
ROADS-GROW

Although the cell variances were roughly proportional to the means, the significance tests below are from analysis of the raw scores; a log transformation yielded similar p values.

Table 2 shows that learning rate increases, of course, with age, $F(3, 144) = 15.45$, $p < .0005$. The poor showing of college *S*s compared to fifth and sixth graders can probably be attributed to the short (two-pair) practice list. The grade school *S*s gained more familiarity with the presentation rate, ITI, and other experimental factors while reaching the practice-list criterion.

Table 2 shows also that except for the reversal in the fifth grade the grammatical lists were learned faster, $F(1, 144) = 5.71$, $p < .025$. The reversal can be attributed to the lone fifth grade *S* who failed to learn before 30 trials; he was assigned to a grammatical list. Although in general the differences between grammatical and ungrammatical lists increased with age, the interaction was not reliable. Tests at each age level showed a significant difference between grammatical and ungrammatical lists for college *S*s, $t(74) = 2.94$, $p < .01$.

EXPERIMENT II

Since grammatical effects did not change reliably with age in Exp. I, younger *S*s were used in this experiment, a variation which required aural presentation of materials. In addition, the design incorporated sentence set as an independent variable.

Method

Subjects.—There were 48 grade school and 48 college *S*s, all with English as their native language. Half the grade school *S*s were first graders, half second graders, from the college-laboratory school. The college *S*s though drawn from the same population had not been in Exp. I. The *S*s from each grade and age level were assigned randomly in blocks of four to one of the experimental lists.

Materials. The eight basic sentences used in this experiment were chosen by the same criterion as in Exp. I and appear in Table 3. They were altered to make two agreement conditions by presence or absence of the suffix /-s/ attached to the noun. Thus no response element ended in /-s/. Further, by adding or not adding the article THE before the noun the sentence Set and No-Set conditions were incorporated. Examples of the four conditions were: THE SHIPS-TALK, SHIPS-TALK, THE SHIP-TALK, and SHIP-TALK.

Four mixed and completely counterbalanced lists were constructed. Each list had two sentences representing each condition, and each sentence appeared in each condition but on a different list.

Each list was typed on memory-drum tape as follows. The eight pairs appeared for a study trial, which was followed by a test trial during which only the stimulus half of the pairs appeared. Pairs were spaced so that with the drum set at a 2-sec. rate each pair appeared for 2 sec. with a 2-sec. IPI and a 10-sec. ITI between both study and test trials. Cues for a ready signal also appeared on the tape in about the middle of each ITI.

TABLE 4

MEAN CORRECT RESPONSES IN EXP. II

Pairs	Age Level			
	Grade School		College	
	Set	No Set	Set	No Set
Grammatical	13.73	13.48	19.27	19.37
Ungrammatical	10.94	11.30	15.94	18.06

Each list was recorded by means of a Bell and Howell magnetic tape recorder. As the materials appeared on the memory drum, *E* spoke them into the recording microphone, with care to space the materials approximately evenly. The materials were spoken with a slight pause between the noun and verb half of the pairs. About 6 sec. before a study trial and upon cue from the drum *E* recorded "Get ready for practice." A similar warning occurred before each test trial. There were three different random orders of both study and test trials; they were repeated continuously to make a total of 12 trials.

A practice list consisting of three pairs, CAT DOG-MAN, HOUSE-CAKE, and SKY-BOX ARM was recorded on a separate reel. The intervals between materials were the same as for the experimental lists and the same ready signals were used. The practice list was preceded on the reel by the recorded instructions for the experiment.

Procedure. The *S* was seated in front of a black screen which hid the tape recorder and *E* from view. The recorded instructions directed him to listen to the materials and to supply the words which were missing (during test trials). Questions from *S* (if any) were answered after the first practice trial. The practice list was continued till *S* reached the learning criterion of one trial correct. Then the practice reel was replaced on the recorder by the reel which presented the experimental list.

Results

Table 4 shows the mean correct responses in each condition over 12 trials for the two age levels. Data from the two age levels were analyzed separately since the variability between grade school *S*s was generally much larger.

It can be seen in Table 4 that for both age levels the grammatical pairs were learned faster. Analysis of the data for grade school *S*s yielded a reliable main Grammatical effect, $F\ (1, 47) = 13.16$, $p < .0001$. Similar results held for college *S*s, $F\ (1, 47) = 15.98$, $p < .0005$.

More interesting in this experiment, however, was the effect of adding the word THE to the noun-verb pairs to enhance sentence set. Table 4 shows that in the grade school sample the Set condition does not differ much from the No-Set condition; the Set effect and its interaction with Grammar were not significant (Fs < 1). But Table 4 shows also that for college *S*s the Set effect was negligible with grammatical pairs but substantial with ungrammatical materials. The Set × Grammar interaction was significant, $F\ (1, 47) = 5.51$, $p < .025$.

Discussion

In both experiments reported here it seemed clear that grammatical habits of noun-verb number agreement can facilitate paired-associate learning. While this outcome is intuitively obvious it does not permit a clear preference for one or another of the current psychological interpretations of language. The present results are consistent with the chunking hypothesis (Miller, 1956) if it is assumed that grammatical pairs are more easily coded, and with the hypothesis of an integrative congruence mechanism (Osgood, 1957). They cannot be explained as simply a contextual reduction in number of alternatives (Miller, Heise, & Lichten, 1951), but they tend to support a dependent probability interpretation if grammatical classes are taken as the sequentially dependent events (Jenkins & Palermo, 1964).

Nevertheless the effects of sentence set in Exp. II lend themselves to an interpretation in the somewhat more conventional language of rote learning. Grammatical word pairs preceded by the article THE were learned at about the same rate as the grammatical pairs not so preceded, which suggests that the facilitative effects of agreement were already maximal for the materials used in the experiment. At the same time the slower learning of ungrammatical pairs under the Set condition by college *S*s seems to require an interference interpretation. The results imply that the operations here called sentence set and grammatical habit refer to somewhat different processes.

Both grade school and college *S*s learned grammatical pairs faster than ungrammatical pairs. Yet when sentencehood was implied under Set conditions but the materials were ungrammatical, the greater interference in learning by college *S*s did not appear in the data of *S*s from the primary grades. This finding could be expected on the assumption that young children have less experience with sentences containing function words. Age differences from grammatical habits might also be expected, but their failure to appear in this experiment may mean that the effects of grammatical number agreement appear earlier in language behavior than the effects of sentence set, and that agreement effects will disappear, if at all, only when *S*s much younger than those in the present experiment are used.

Taken together the results of these experiments show a similarity between grammatical phenomena and the behavior ordinarily studied in the rote-learning laboratory, and they tend to affirm the relevance of associative theory to the problems of accounting for language.

REFERENCES

CHOMSKY, N. *Syntactic structures.* 's-Gravenhage, Holland: Mouton, 1957.

JENKINS, J. J., & PALERMO, D. S. Mediation processes and the acquisition of linguistic structure. *Monogr. Soc. Res. Child Develpm.*, 1964, **29**(1, Ser. No. 92), 141-168.

LASHLEY, K. S. The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior.* New York: Wiley, 1951. Pp. 112-135.

MILLER, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.*, 1956, **63**, 81-97.

MILLER, G. A., GALANTER, E., & PRIBRAM, K. H. *Plans and the structure of behavior.* New York: Holt-Dryden, 1960.

MILLER, G. A., HEISE, G. A., & LICHTEN, W. The intelligibility of speech as a function of the context of the test materials. *J. exp. Psychol.*, 1951, **41**, 329-335.

OSGOOD, C. E. A behavioristic analysis of perception and meaning as cognitive phenomena. In J. S. Bruner et al., *Contemporary approaches to cognition.* Cambridge: Harvard Univer. Press, 1957.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, Bureau of Publications, 1944.

(Received July 13, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 6, 575-579

EFFECTS OF NUMBER AND PERCENTAGE OF REWARDED TRIALS ON THE ACQUISITION AND EXTINCTION OF LEVER PRESSING USING A DISCRETE-TRIAL PROCEDURE[1]

JOHN J. PORTER AND JAMES J. HUG

University of Wisconsin, Milwaukee

8 groups of 10 hooded rats received 12, 32, 64, or 96 acquisition trials with 50% or 100% water reward. 16 bar presses defined a trial. Starting and total speeds were examined as were trials-to-extinction criteria of 1, 3, 5, or 10 min. Acquisition performance was dependent on the interaction of number of training trials with percentage of reward, and also an increasing function of number of training trials. Within-Ss comparisons showed that the 50% Ss responded significantly faster following nonreward than reward. The 100% reward groups showed little change in resistance to extinction after 32 prior training trials, and no less resistance to extinction than the 50% Ss.

In the past few years it has become increasingly evident that the pattern of responding obtained during free-operant extinction is quite unlike the response patterns in discrete-trial runway extinction. In free-operant conditioning, Dyal and Holland (1963), Miles (1956), Perin (1942), and Williams (1938) all found resistance to extinction to be an increasing function of the number of acquisition trials on 100% reward, while Denny, Wells, and Maatsch (1957) obtained similar results using 4:1 fixed-ratio reinforcement. In a recent runway study, using a 30-sec. intertrial interval (ITI), Hill and Spear (1963) found a positive relationship between number of training trials and resistance to extinction early in extinction, but no evidence of this relationship in later extinction. Other straight-alley investigations by Bacon (1962), Ison (1962), North and Stimmel (1960), and Siegel and Wagner (1963) have found an inverse relationship between number of acquisition trials and resistance to extinction.

As Amsel, MacKinnon, Rashotte, and Surridge (1964) have pointed out, the contrast between the results of free-operant and discrete-trial procedures suggests the need for experimental comparisons of these dissimilar response situations. The present study used a discrete-trial, lever-pressing response as an analog to the runway CR, in order to effect such a comparison. Each response chain of 16 bar presses was followed by a 30-sec. ITI, this sequence constituting a trial. Either 12, 32, 64, or 96 such acquisition trials were administered, followed by 40 extinction trials. For half the *S*s reinforcement occurred on a randomly selected 50% of the acquisition trials; the other *S*s received continuous reinforcement.

METHOD

Subjects.—Ninety (54 male and 36 female) hooded rats, 95–135 days old, served as *S*s.

Apparatus.—The *S*s were run in a 7×12×8 in. ventilated and sound-insulated chamber, illuminated by a small 7½-w. bulb. A retractable response lever and liquid dipper were mounted on one wall of the chamber. Beginning with the insertion of the bar into the box, the latency of the first bar press and

---

[1] Supported by National Science Foundation Grant G-22832.

the total time from bar insertion to the last response were recorded.

Design.—The *Ss* were assigned to a 2×4 factorial design of 12, 32, 64, or 96 acquisition trials under partial (50%) or continuous reward. On the average, for the 50% *Ss*, in a block of 32 trials a nonreward (NR) trial was followed by a rewarded (R) trial 12 times, while an NR trial was followed by an NR trial, or an R trial by an R trial 10 times. A maximum of 32 acquisition trials was run daily; on the day following acquisition, all *Ss* began 2 days of extinction to a daily criterion of 10 min. without completing a trial. A maximum of 28 extinction trials was permitted on Day 1, and 12 trials on Day 2. A control group of 10 *Ss* had the same preliminary training as the other groups and was tested for resistance to extinction on the following day, without any acquisition training.

Procedure.—Preliminary training required 13 days; on the first 9 days drinking time was reduced from 60 to 20 min. in 5-min. daily decrements. On Day 10 (magazine training) *Ss* were given thirty .06-ml. water rewards, the amount used throughout the experiment. On Day 11, each *S* received 30 rewards while learning to bar press. On Day 12, *Ss* made five rewarded bar presses, followed by three reinforced discrete trials; on these three trials, as between all trials thereafter, the chamber light was out and the bar retracted during the 30-sec. ITI. On Day 13, *Ss* received 21 discrete trials, using increasing reinforcement ratios from 2:1 to 14:1. The ratios, and thus the number of responses required for reward, increased by two after every third trial. On all preliminary and training trials, rewards were delivered at the end of a response chain. On Day 14, acquisition training began for the experimental groups, while the control group began extinction. During all sessions *Ss* were left in the test chamber until they had completed their daily trials, then immediately returned to their home cages and had water available for 20 min. Food was always available in the home cages.

RESULTS

While the starting and total-speed results were similar, group differences were consistently larger for total speeds; consequently, the data presented here concern only total speeds.

Acquisition. Response speeds during the last three blocks of four trials, for each experimental group, were an increasing function of the number of training trials; $F\ (3,\ 72) = 38.39$, $p < .001$. However, the overall effect of percentage of reinforcement was not significant, apparently due to the early superiority of the 100% *Ss* over the 50% *Ss*; a superiority whose later reversal was reflected in a significant Trials × Percentage of Reward interaction; $F\ (3,\ 72) = 2.92$, $p < .05$. In agreement with other investigators (Goodrich, 1959; Wagner, 1961; Weinstock, 1958), the 50% *Ss* ran slower early in training, Trials 21–32, $t\ (18) = 2.16$, $p < .05$, and faster late in training, Trials 85–96, $t\ (18) = 2.76$, $p < .02$. However, no reliable performance difference prevailed between the 100% and 50% groups over Trials 1–12 or 53–64.

Figure 1 presents response speeds for the two 96 trial groups. The 12, 32, and 64 trial groups were not presented in Fig. 1 because they did not differ significantly from the 96 groups at any point in training. In order to determine if there were any reliable differences in response speed following R and NR trials, the speeds of the 96–50 group were separated into post-R and post-NR speeds. The two light lines in Fig. 1 represent these speeds in blocks of four R or NR trials. The two heavy lines represent the overall speed of the 96–100 and 96–50 groups, blocked in units of eight trials. As is apparent from Fig. 1, all *Ss* consistently ran fastest after an NR trial. When analyzed over all 12 blocks of post-R and post-NR trials, this difference in speed was highly reliable; $t\ (9) = 8.60$, $p < .01$. The response speeds of the 100% *Ss* were evaluated separately against the post-R and post-NR speeds of the 50% *Ss*,

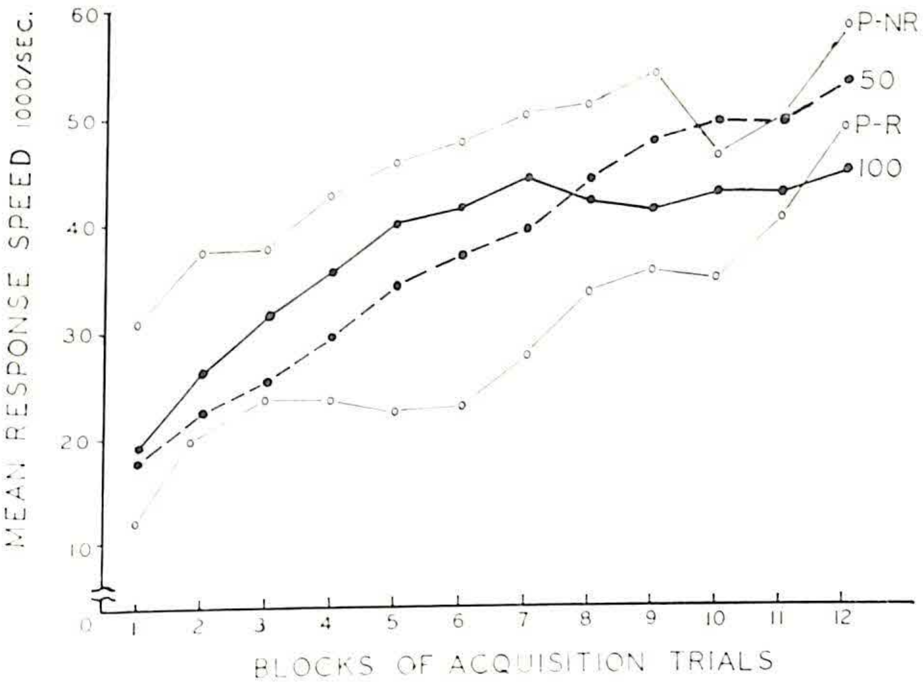


FIG. 1. Mean response speed for the 96–100 and 96–50 groups (the heavy lines) blocked in units of eight acquisition trials. (The two light lines represent the within-Ss response speeds, in blocks of four trials, after nonreinforced, P-NR, and after reinforced, P-R, trials.)

using the initial three and final three blocks of four acquisition trials. It was found that the 100% *Ss* ran significantly slower than the 50% *Ss* after NR trials, on both initial and final blocks of acquisition trials; t (18) = 4.48, 2.28; $p < .01$ and .05, respectively; however, the speed of the 100% group did not differ significantly from that of the 50% group after R trials, for either the initial or final blocks of acquisition training.

Extinction speeds.—The effects of percentage of reward, number of training trials, and blocks of extinction trials were analyzed by means of a three-dimensional mixed design. Response speed was an increasing function of number of prior acquisition trials, F (3, 72) = 50.53, $p < .001$, but not of prior 50% or 100% reward. The decrement in extinction speed over the first four blocks of trials was significant, F (3, 216) = 128.70, $p < .001$; however, the only significant interaction was between acquisition trials and percentage of reward, F (3, 72) = 6.49, $p < .001$, leading to an examination of the simple effects of these two variables. For the 96 trial groups, 50% reward led to significantly slower extinction than 100% reward, t (18) = 3.35, $p < .01$, while the converse was true for the 32 trial groups, t (18) = 2.61, $p < .05$. No significant differences were found between extinction speeds of the 100% and 50% *Ss* in the 12 and 64 trial groups. The control group was significantly slower than the slowest experimental group, Group 12–50, on both extinction speed, t (18) = 7.78, $p < .01$, and trials to extinction, t (18) = 5.17, $p < .01$.

Trials to extinction.—Figure 2 displays the number of trials to extinction plotted in terms of 10-, 5-, 3-, and 1-min. extinction criteria. Using a more stringent criterion merely

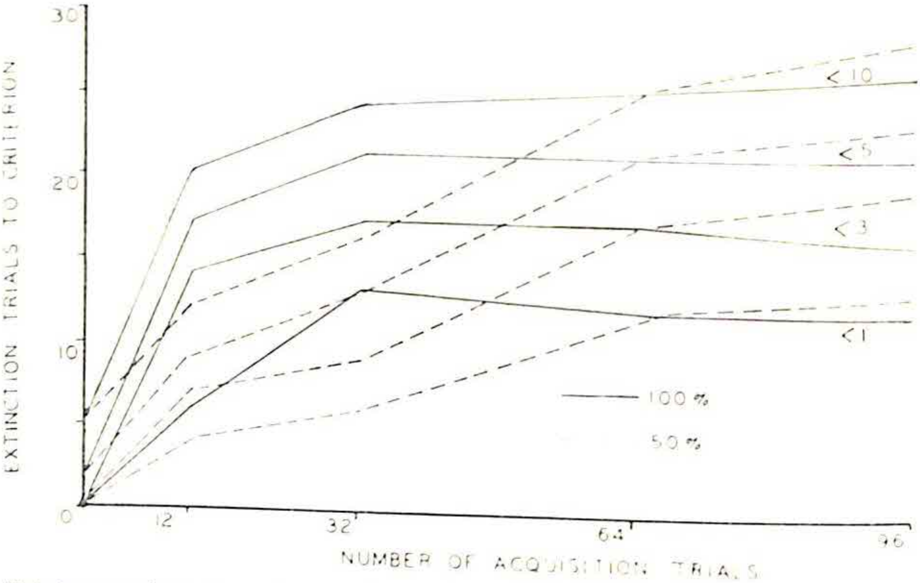


FIG. 2. Trials to extinction under the 10-, 5-, 3-, and 1-min. extinction criteria as a function of the number of prior acquisition trials at either 50% or 100% reinforcement.

lowered the extinction gradient without changing its form. Since the outcome of the four separate analyses of variance for each criterion were essentially the same, only the results from the 10-min. criterion are discussed. Number of training trials, F (3, 72) = 24.46, $p < .001$; percentage of reward, F (1, 72) = 10.78, $p < .005$; and the interaction of these variables, F (3, 72) = 6.05, $p < .005$, were all highly reliable determiners of extinction performance. From Fig. 2 it was apparent that, through 64 trials, 100% reward yielded greater resistance to extinction than 50% reward. Statistically, this finding was confirmed after 12, t (18) = 5.59, $p < .01$, and 32, t (18) = 2.91, $p < .01$, acquisition trials, but not after 64 or 96 trials. Closer examination of the extinction gradient of the 100% *S*s, using the Newman-Keuls test (Winer, 1962), showed no significant change in trials to extinction after 32 acquisition trials, a result similar to the late extinction findings of Hill & Spear (1963). In contrast to the 100% *S*s, the 50% *S*s showed an orderly increase in trials to extinction from 12 to 96 prior acquisition trials.

DISCUSSION

In the present study the large frustration effect, shown by the significant superiority of the 96–50 group over the 96–100 group near the end of acquisition, resembled the runway findings of Goodrich (1959), Wagner (1961), and Weinstock (1958). However, unlike these runway studies, the faster speeds of the 50% *S*s were due to a gain in total speed after R trials, rather than any change in speed after NR trials. In fact, the persistent difference in speed after NR and R trials was similar to the effects of frustration on second alley response speed in the double runway used by Amsel and Roussel (1952) and others. The late increase in speed following R trials may have been due to the generalization of frustration from post-NR trials to post-R trials.

Turning to extinction, the gradients obtained in this study after 100% reward resemble those of free operant conditioning more than those of runway experiments. Certainly there was no evidence of decreased resistance to extinction with overlearning, as has been observed in runways under conditions of 100% reward. On the other hand, after large amounts of training on 100% reward, the increase in resistance to extinction was certainly much less than is typical of free operant conditioning. Further, the lack of the usual partial-reinforcement superiority in trials to extinction is contrary to the usual findings in both free-operant and discrete-trial extinction.

While the present results do not account for the difference in resistance to extinction observed in free-operant and discrete-trial responding, they do point out that the introduction of a discrete trial was not sufficient to change bar-pressing behavior to that expected in the discrete-trial runway. Possibly the relatively extensive pretraining, necessary to get *S* to drink, bar press, and acquire the chained response, provide sufficient partial reinforcement that the introduction of partial reinforcement of *the chained response*, during acquisition, had little effect on extinction behavior. It would seem that an investigation of the effects of varying amounts of pretraining of the bar press, and of a response more easily trained than the bar press, is in order.

REFERENCES

AMSEL, A., MACKINNON, J. R., RASHOTTE, M. E., & SURRIDGE, C. T. Partial reinforcement (acquisition) effects within subjects. *J. exp. Anal. Behav.*, 1964, **7**, 135–138.

AMSEL, A., & ROUSSEL, J. Motivational properties of frustration: I. Effect on a running response of the addition of frustration to the motivational complex. *J. exp. Psychol.*, 1952, **43**, 363–368.

BACON, W. E. Partial-reinforcement extinction effect following different amounts of training. *J. comp. physiol. Psychol.*, 1962, **55**, 998–1003.

DENNY, M. R., WELLS, R. H., & MAATSCH, J. L. Resistance to extinction as a function of the discrimination habit established during fixed-ratio reinforcement. *J. exp. Psychol.*, 1957, **54**, 450–456.

DYAL, J. A., & HOLLAND, T. A. Resistance to extinction as a function of number of reinforcements. *Amer. J. Psychol.*, 1963, **76**, 332–333.

GOODRICH, K. P. Performance in different segments of an instrumental response chain as a function of reinforcement schedule. *J. exp. Psychol.*, 1959, **57**, 57–64.

HILL, W. F., & SPEAR, N. E. Extinction in a runway as a function of acquisition level and reinforcement percentage. *J. exp. Psychol.*, 1963, **65**, 495–500.

ISON, J. R. Experimental extinction as a function of number of reinforcements. *J. exp. Psychol.*, 1962, **64**, 314–317.

MILES, R. C. The relative effectiveness of secondary reinforcers throughout deprivation and habit-strength parameters. *J. comp. physiol. Psychol.*, 1956, **49**, 126–130.

NORTH, A. J., & STIMMEL, D. T. Extinction of an instrumental response following a large number of reinforcements. *Psychol. Rep.*, 1960, **6**, 227–234.

PERIN, C. T. Behavior potentiality as a joint function of the amount of training and the degree of hunger at the time of extinction. *J. exp. Psychol.*, 1942, **30**, 93–113.

SIEGEL, S., & WAGNER, A. R. Extended acquisition training and resistance to extinction. *J. exp. Psychol.*, 1963, **66**, 308–310.

WAGNER, A. R. Effects of amount and percentage of reinforcement and number of acquisition trials on conditioning and extinction. *J. exp. Psychol.*, 1961, **62**, 234–242.

WEINSTOCK, S. Acquisition and extinction of a partially reinforced running response at a 24-hr. intertrial interval. *J. exp. Psychol.*, 1958, **56**, 151–159.

WILLIAMS, S. B. Resistance to extinction as a function of the number of reinforcements. *J. exp. Psychol.*, 1938, **23**, 506–522.

WINER, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

(Received July 14, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 6, 580–584

TEST FOR A LEARNED DRIVE BASED ON THE HUNGER DRIVE[1]

JOHN H. WRIGHT[2]

University of Virginia

On 24 occasions over a 36-day period, 42 male albino rats were placed in distinctive boxes for 1 hr. when 1 hr. hungry and in qualitatively different boxes for 1 hr. when approximately 20 hr. hungry. 14 Ss were assigned to each of 3 conditions of delay of return to home cage and food following strong-drive training: (a) immediate return to home cage and food, (b) immediate return to home cage but with food delayed, and (c) delayed return to both home cage and food. Later, Ss were given a 1-hr. eating period in each box under 11-hr. hunger. Significantly more food was eaten in the strong-drive than in the weak-drive box by all groups with Ss of Cond. b tending to eat more in the strong-drive box and less in the weak-drive box than Ss of Cond. a and c.

Many experiments have shown that learned drive is established when a neutral stimulus is repeatedly paired with the onset of a primary drive state. In practically all of the attempts to establish learned drive, however, the procedure has been to pair a neutral stimulus with the onset of a primary aversive drive state such as pain produced by electric shock (e.g., Brown & Jacobs, 1949; May, 1948; Miller, 1948). To extend the generality of the concept of learned drive, investigators have attempted to demonstrate the motivating efficacy of a stimulus paired with the onset of a primary appetitive drive state such as hunger or thirst. However, most of the attempts to demonstrate learned drive in this manner have been unsuccessful (e.g., Howard & Young, 1962; Myers & Miller, 1954; Novin & Miller, 1962; Siegel & MacDonnell, 1954).

One of the few studies to report positive evidence for learned appetitive drive is that of Calvin, Bicknell, and Sperling (1953). These investigators placed rats in a distinctive box for $\frac{1}{2}$ hr. daily, one group under weak hunger and a second group under strong hunger. When later placed in the box under an intermediate level of primary hunger and given food for the first time, the strong-drive group ate slightly but significantly more than the weak-drive group. The authors attributed this difference in intake to the conditioning of hunger to the cues of the box in the case of the animals trained under strong hunger. A replication of this study by Siegel and MacDonnell (1954), however, failed to confirm these results.

The experiment to be reported represents a further attempt to demonstrate a learned drive based on hunger. The procedure employed was similar to that of Calvin et al. (1953) and Siegel and MacDonnell (1954). In the present study, however, each *S* was placed in a distinctive box under weak hunger and in a qualitatively different box under strong hunger and was later tested in both boxes under an intermediate level of primary hunger. During training *S*s were maintained on cyclic food deprivation and were placed daily in the drive-training boxes under the appropriate level of hunger. Since daily feeding necessarily followed closely strong-drive training, and could be a possible factor responsible for greater intake in the strong-drive than in the weak-drive box, *S*s were assigned to three conditions of delay of return to the home cage and food following strong-drive training: (*a*) immediate return to home cage and food, (*b*) immediate return to home cage but with food delayed, and (*c*) delay in a neutral box before return to home cage and food.

---

[1] This investigation was supported in part by a Public Health Service Grant (No. MH-04920) to Frank W. Finger and L. Starling Reid and is based on a dissertation submitted to the Graduate Faculty of the University of Virginia in partial fulfillment of the requirements for the PhD degree. The author wishes to acknowledge the assistance throughout the course of the investigation of L. Starling Reid. A preliminary report of the results of this study was presented at the September 2, 1963 meeting of the American Psychological Association.

[2] Now at Wake Forest College.

METHOD

Subjects.—The *S*s consisted of 42 experimentally naive, male albino rats of the Wistar strain. They ranged from 83 through 96 days of age at the beginning of the habituation phase of the experiment.

Apparatus and general environmental conditions.—The *S*s, housed individually in 7×7×10 in. home cages throughout the experiment, were located in a small, relatively quiet room in which the temperature varied from 80 through 86° F. No attempt was made to control or to record humidity. The light-dark cycle was automatically controlled and consisted of two 12-hr. periods.

Forty-two wooden boxes (inside dimensions 7×7×10 in.) were used as the drive-training boxes. A hole in the lower front end of each box constituted the major source of illumination and ventilation; through this hole was inserted the mouth of a drinking tube during testing. The interior of 21 of the boxes was painted flat black, the remaining 21 flat white. The interior of the wooden delay boxes (inside dimensions 6 × 6 × 9 in.) was unpainted.

Procedure.—The two major independent variables were strength of drive during training and delay of return to the home cage after drive training. The *S*s were maintained throughout the experiment on a liquid diet described by Teitelbaum and Campbell (1958). Following a 10-day period of adjustment to a 22-hr. cyclic food-deprivation schedule, a 36-day period of training was begun. During this period *S*s received in random order weak-drive training on 12 of the days, strong-drive training on 12 of the days, and both weak- and strong-drive training on 12 of the days. Drive training consisted of placing *S* in a box under the appropriate level of hunger for a period of 1 hr. Thus, over the 36 days of training all *S*s received 24 weak-drive and 24 strong-drive training "trials."

Fourteen *S*s were assigned at random to each of the three conditions of delay of return to the home cage and food following strong-drive training. These conditions were: Group I—immediate return to home cage and food after strong-drive training; Group II—immediate return to the home cage after strong-drive training but with food delayed ½ hr.; and Group III—delayed ½ hr. in a neutral box before return to home cage and food after strong-drive training. Because no food followed weak-drive training, Groups I and II were treated identically at this time, both being returned to the home cage. Group III was delayed ½ hr. before return to the home cage after weak-drive training.

Half the *S*s in each delay group received weak-drive training in a black box and strong-drive training in a white box. For the remaining half of each group the opposite was the case. All groups began weak-drive training under 1 hr. of hunger. Groups II and III began strong-drive training under 20.5 hr. of hunger and Group I under 21 hr. of hunger.

On the day following the last day of training *S*s were fed at the usual time. Eleven hours following this eating period the first of the two 1-hr. tests was begun. The tests consisted of permitting *S* to eat in the weak- and strong-drive boxes for the first time. Ten hours following Test 1 daily feeding was again given at the usual time. Twenty-two hours later *S*s were fed as usual and 11 hr. later the second 1-hr. test was begun.

For each delay condition, box color was counterbalanced during testing. Half of each group was tested first in a black box and second in a white box; the remaining half of each group was tested first in a white box and second in a black box. Within the restrictions imposed by the counterbalancing of box color during testing, the box in which weak and strong hunger had been experienced was also counterbalanced during testing.

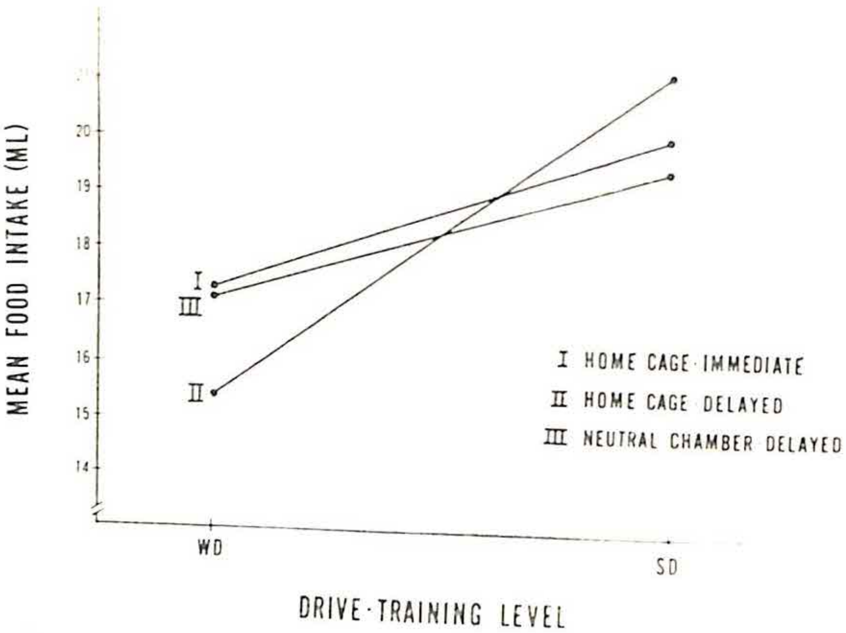


FIG. 1. Mean food intake for the three delay groups tested in the weak-drive and strong-drive boxes.

Thus, eight *S*s in each delay group were tested first in the box in which strong hunger had been experienced and second in the box in which weak hunger had been experienced. The remaining six *S*s of each delay group were tested first in the weak-drive box and second in the strong-drive box.

RESULTS

The results of the two tests separately present a somewhat different picture. However, these differences are for the most part attributable to factors which were counterbalanced over the two tests. For this reason only the intake data for both tests combined will be considered.

Mean 1-hr. food intake for the three delay groups tested in the weak- and strong-drive boxes is presented in Fig. 1. As was the case for each of the two tests considered separately, strong-drive training led to significantly more intake during testing than weak-drive training, $F\ (1, 30) = 55.00$, $p < .001$. Furthermore, this effect was significant for all conditions of delay, $p < .025$. Thus, all groups ate reliably more in the strong-drive than in the weak-drive boxes.

Although not significant as a main effect the delay variable interacted significantly with drive-training level, $F\ (2, 30) = 5.52$, $p < .01$. That this interaction was mainly due to the difference of Group II from Groups I and III is suggested by individual comparisons, which revealed a tendency for Group II to eat more than Groups I and III during the strong-drive test but less than these groups during the weak-drive test.

DISCUSSION

The finding of significantly greater intake in the strong-drive than in the weak-drive boxes for all delay groups seems to demonstrate energizing effects on consummatory behavior of cues previously paired with hunger. These results provide evidence for conditioned hunger and are in agreement with the findings of Calvin et al. (1953).

A possible reason for the failure of previous studies to demonstrate learned appetitive drive is the fact that the onset of most appetitively induced drive states occurs too gradually to enable a stimulus to be paired effectively with a rapid increment in drive. Fromer (1962) obtained only weak conditioned fear when the onset of a painful electric shock paired with a neutral stimulus was very gradual, presumably paralleling the onset of an appetitive drive such as hunger. This problem has led investigators (e.g., Andersson & Larsson, 1956; Novin & Miller, 1962) to search for experimental procedures which induce appetitive drive more rapidly than deprivation, but such attempts have not resulted in the successful demonstration of acquired appetitive drive.

The use of a within-*S* experimental design may account for why positive evidence for conditioned hunger was found in the present experiment. In addition to the greater statistical precision of within-*S* designs, it has been shown that larger absolute performance differences occur between experimental conditions administered to the same *S*s than when these conditions are administered to separate groups of *S*s (Grice & Hunter, 1964).

It is obvious, however, that conditioned hunger cannot account for the significant interaction of drive-training level with delay of return to home cage and food. Furthermore, this interaction cannot be accounted for by assuming that the close temporal proximity of strong-drive training to daily feeding led to the conditioning of r_g to the cues of the strong-drive box. On the basis of r_g theory (Spence, 1956) one would expect greater conditioning of r_g and, consequently, greater strong-drive-box intake by Group I than by Groups II and III because for Group I food immediately followed strong-drive training. As can be seen from Fig. 1, however, the strong-drive-box intake of Group I failed to differ from that of Group III and, in addition, was actually less than that of Group II.

On the other hand, the procedure followed for Group II resembles quite strikingly procedures which have been employed in experimental studies of frustration. The *S*s of Group II, because of their return to the home cage in the absence of food following training under strong drive, were placed in the presence of stimuli to which the primary eating response is strongly associated. The absence of food in this situation could well have produced primary frustration. Many studies (e.g., Amsel & Hancock, 1957; Amsel & Roussel, 1952; Amsel & Ward, 1954) have shown that primary frustration is produced when food reward is withheld from a hungry rat in a situation in which he has previously been consistently rewarded. It seems reasonable that the conditioning of frustration to the cues of the strong-drive box in the case of Group II could account for the greater strong-drive intake by this group than by Groups I and III. On the other hand, it should be noted that the presence of frustration conditioned to the cues of the strong-drive, but not the weak-drive box, could also account for the overall difference of strong- vs. weak-drive training and, thus, obviate the need to interpret the present results in terms of conditioned hunger. However, a source of conditioned frustration, clearly present in the case of Group II, cannot readily be identified for Groups I and III.

The only noteworthy point concerning the ordering of groups for the weak-drive test is the smaller intake of Group II as compared with Groups I and III, and this may also be accounted for in terms of a frustration hypothesis. The elicitation of conditioned frustration by the cues of the strong-drive box in the case of Group II could well have provided the basis for a superior discrimination between the weak- and strong-drive boxes on the part of this group. Although evidence on this point is meager, Amsel and Ward (1954) have presented findings which suggest that frustration-produced stimuli may serve as discriminative stimuli.

It is clear that the results of the present experiment cannot be accounted for solely in terms of conditioned hunger.

While the overall difference of strong- vs. weak-drive training may be interpreted best in terms of conditioned hunger, a conditioned-frustration hypothesis is needed to account for the difference of Group II from Groups I and III.

REFERENCES

AMSEL, A., & HANCOCK, W. Motivational properties of frustration: III. Relation of frustration effect to antedating goal factors. *J. exp. Psychol.*, 1957, **53**, 126–131.

AMSEL, A., & ROUSSEL, J. Motivational properties of frustration: I. Effect on a running response of the addition of frustration to the motivational complex. *J. exp. Psychol.*, 1952, **43**, 363–368.

AMSEL, A., & WARD, J. S. Motivational properties of frustration: II. Frustration drive stimulus and frustration reduction in selective learning. *J. exp. Psychol.*, 1954, **48**, 37–47.

ANDERSSON, B., & LARSSON, S. An attempt to condition hypothalamic polydipsia. *Acta physiol. Scand., Stockholm*, 1956, **36**, 377–382.

BROWN, J. S., & JACOBS, A. The role of fear in the motivation and acquisition of responses. *J. exp. Psychol.*, 1949, **39**, 747–759.

CALVIN, J. S., BICKNELL, E. A., & SPERLING, D. S. Establishment of a conditioned drive based on the hunger drive. *J. comp. physiol. Psychol.*, 1953, **46**, 173–175.

FROMER, R. The effect of several shock patterns on the acquisition of the secondary drive of fear. *J. comp. physiol. Psychol.*, 1962, **55**, 142–144.

GRICE, G. R., & HUNTER, J. J. Stimulus intensity effects depend upon the type of experimental design. *Psychol. Rev.*, 1964, **71**, 247–256.

HOWARD, T. C., & YOUNG, F. A. Conditioned hunger and secondary rewards in monkeys. *J. comp. physiol. Psychol.*, 1962, **55**, 392–397.

MAY, M. A. Experimentally acquired drives. *J. exp. Psychol.*, 1948, **38**, 66–77.

MILLER, N. E. Studies of fear as an acquirable drive: I. Fear as motivation and fear reduction as reinforcement in the learning of new responses. *J. exp. Psychol.*, 1948, **38**, 89–101.

MYERS, A. K., & MILLER, N. E. Failure to find a learned drive based on hunger; evidence for learning motivated by "exploration." *J. comp. physiol. Psychol.*, 1954, **47**, 428–436.

NOVIN, D., & MILLER, N. E. Failure to condition thirst induced by feeding dry food to hungry rats. *J. comp. physiol. Psychol.*, 1962, **55**, 373–374.

SIEGEL, P. S., & MACDONNELL, M. F. A repetition of the Calvin-Bicknell-Sperling study of conditioned drive. *J. comp. physiol. Psychol.*, 1954, **47**, 250–252.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

TEITELBAUM, P., & CAMPBELL, B. A. Ingestion patterns in hyperphagic and normal rats. *J. comp. physiol. Psychol.*, 1958, **51**, 135–141.

(Received July 15, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 6, 585-597

INTRASESSION ADAPTATION AND INTERSESSION EXTINCTION OF THE COMPONENTS OF THE ORIENTING RESPONSE[1]

CHARLES R. GALBRECHT, ROSCOE A. DYKMAN, WILLIAM G. REESE, AND TETSUKO SUZUKI

University of Arkansas Medical Center

Intrasession (adaptation) and intersession (extinction) changes in the skin resistance (SR), heart rate (HR), respiratory rate (RR), and muscle action potentials (MP) components of the orienting response (OR) to auditory stimuli of moderate intensity were compared. Adaptation and extinction curves were very similar in SR and similar in MP. There was essentially no evidence of adaptation or extinction in HR and RR or indeed systematic responsiveness. These findings support the logical inference: Adaptation and extinction are equivalent processes mediated by the same neural mechanisms. Significant bilateral differences in prestimulus levels of SR did not affect responses; in each system Ss maintained their rank-order position in the group across sessions; the responses studied provided little evidence of startle components; and the generally larger reactions in SR and MP on the 1st tone each day were not related to arousal levels as measured by prestimulus levels of functioning.

Research on the orienting response (OR) has been greatly stimulated by the work of Sokolov (1960, 1963). Sokolov (1960) defines the OR as a response produced by any novelty in the environment. He says the OR has three general properties: It is nonspecific as regards the modality and the intensity of the stimulus and is subject to habituation or extinction. The OR, according to Sokolov, may be discriminated from "two other general types of unconditioned reflexes—the adaptive reflex and the defensive reflex [p. 189]." The normal adaptive reaction to warmth, e.g., is vasodilation in both head and hand. But the first few presentations of warmth dilate the blood vessels of the head and constrict those of the finger, the latter pattern being presumably the common effects of a great variety of novel stimuli.

Sokolov has studied only some components of the OR, and it remains to be seen whether different orienting stimuli do indeed have the common effects he assumes. We predict, e.g., that studies involving a reasonable sample of bodily functions will reveal important differences between ORs to visual and auditory stimuli. We certainly must discriminate general bodily effects such as changes in heart rate (HR), respiratory rate (RR), skin resistance (SR), and muscle action potentials (MP) from more specific or localized bodily effects such as gastrointestinal secretions to a specific food, flexion to shock, vasodilation to warmth, and motor-orienting reactions. The best way to do this—indeed the only way—is to attempt to classify reactions, as Sokolov does, on the basis of their

---

[1] This investigation was supported in part by Grant MH-01091 and continuations thereof from the National Institute of Mental Health, United States Public Health Service, Bethesda, Maryland. The work of R. A. Dykman was supported by a career development award (MH-2504) from the National Institute of Mental Health. In addition, a corollary grant to MH-2504 (M-6009) was used to support this work.

generality or specificity with respect to stimuli.

From our point of view, the OR is a gross-organismic response elicited by any stimulus of the CS class which excludes all incentive agents and noxious stimuli. The OR includes only those components that appear reliably the first time or two the stimulus is presented. But the components vary within and between species; e.g., nonreinforced tones increase HR in most dogs but rarely do in humans (Dykman & Gantt, 1959; Dykman, Reese, Galbrecht, & Thomasson, 1959; Robinson & Gantt, 1947). Further, the OR has both specific components—stimulus-directed movements such as eyelid changes and motor orientation of the head—and nonspecific components—general autonomic reactions and changes in tone in muscles not used in orienting to signals.

Some question may arise as to whether reactions to simple stimuli presented on a fixed time sequence, as in the present experiment, qualify as ORs. Pavlov (1927, p. 29) would classify such reactions as ORs, Sokolov would, and we do. Further, to view reactions to simple stimuli as ORs, relates relatively insignificant units of behavior to larger units of behavior, curiosity, and exploration.

A number of American investigators have studied the OR, or the equivalent reaction under other names (Darrow, 1929). The literature reviewed here will be confined to researches most pertinent to the present investigation. At least two studies indicate that the GSR component of the OR habituates by the fifth trial (Bitterman & Holtzman, 1952; Howe, 1958). Spontaneous recovery of the GSR component has also been shown in a test-retest situation with faster habituation on retest 6–8 wk. later (Rachman, 1960). Davis, Buchwald, and Frankmann (1955), in one of the best multiple-component studies ever done on reactions to auditory signals, reported on the effects of trials and intensity of stimuli. Davis et al. (1955) and Davis (1953) show habituation gradients for MP. The MP component had a short latency, 200–300 msec., suggesting that a startle reflex preceded the OR.

Dykman et al. (1959) provided information concerning GSR, RR, and HR components of the OR in humans, the GSR component being most consistent. The GSR clearly decreased in magnitude but did not completely extinguish in a series of 18 tones spaced at 1-min. intervals.

Allen, Hill, and Wickens (1963) investigated the OR as a "formation of temporal relations among stimuli, with GSR the measure [p. 309]." Corah and Stern (1963) reported adaptation of the GSR component of the OR in children with a more rapid habituation in the second daily session than the first. They also reported habituation of spontaneous activity in SR (nonstimulus coupled oscillations) which neither we (Wilson & Dykman, 1960), Lacey and Lacey (1958a), nor Johnson (1963) found. Unger (1964) studied the vasoconstrictive component of the OR to acoustically novel stimuli and found evidence of adaptation within a single session in most *S*s. Some investigators attempt to maintain a high level of novelty by variation of stimulus parameters. By contrast, this study is restricted to a single tone presented at fixed intervals with the aim of minimizing the number of variables involved in adaptation or extinction.

The present report represents a continuation of our studies of the OR and provides information on (*a*) intra-

and intersession gradients and their interrelationship; (*b*) the relation of responses to arousal or system activation as defined by prestimulus levels of functioning; (*c*) relationships between different techniques of scoring physiological responses; (*d*) bilateral differences in SR and MP; (*e*) startle components of the OR as reflected by MP; and (*f*) adaptation of physiological background activity.

METHOD

Subjects.—The *Ss* were 20 male freshman medical students between 20 and 22 yr. of age. They were given no information concerning the nature of the experiment other than that the study involved eight daily sessions and that they would be paid $2.00 per session. The *Ss* were instructed not to discuss the procedures among themselves.

Apparatus.—The instrumentation has been described in detail elsewhere (Dykman, Ackerman, Galbrecht, & Reese, 1963). In brief, SR was recorded by two Fels dermohmmeters and two Esterline-Angus recorders; HR by a Fels cardiotachometer and Epsco galvanometer; RR by a Statham strain gauge, Epsco carrier preamplifier, and Epsco galvanometer; and MP by Epsco integrator preamplifiers and galvanometers. The Epsco preamplifiers sum the MP occurring in each .3-sec. interval, the height of the tracing being a relative measure of the total activity during the period. Instrumentation errors which could produce systematic differences in SR levels on the two sides of the body were controlled by a random alternation of dermohmmeters among *Ss* and by a careful balancing of the meters.

Procedure.—Upon arrival at the laboratory, *S* was told to wash his feet, this to insure a clean site for the SR electrodes. He was then seated in an upholstered chair with his feet resting on an ottoman, the chair located in a sound-shielded room isolating him from *Es* and the recording apparatus. Two SR electrodes (cup-type with zinc plate and filled with zinc sulphate) were fastened to each leg, one on the inferior surface of the plantar arch and the other just superior to the internal malleolus. Standard HR electrodes saturated with Sanborn jelly were attached to the two wrists, or, on occasion, to the left wrist and right leg, depending upon the lead giving the largest R wave. A pneumograph was placed around the chest or abdomen at the point of maximum expansion. The MP was recorded from two EEG electrodes on each arm, 2 in apart on the midpoint of the right and left forearms over the flexor carpi radialis and flexor digitorum sublimus muscles. A set of earphones cushioned by sponge rubber was placed on *S*'s head and he was left alone in the room.

The *S* was instructed over a loudspeaker to stay awake and to keep his head and eyes turned toward the front. Basal recordings were then made for 5 min., followed by a series of 12 tones, each 800 cps, 60 db., and 5 sec. in duration. The intertone interval was 1 min. ± 5 sec., the variations enabling the presentation of tones at the peak of inspiration. An experimental session terminated with a stimulus-free period of 3 min. It was not possible to run *Ss* for 8 consecutive days, and intersession intervals varied from 2 to 3 days. The procedures for each session were identical.

Physiological measures.—Prestimulus levels of functioning were measured as follows: for SR, the exact level in ohms at the onset of each tone; for HR, the shortest interval between any two consecutive R waves during the 5-sec period just preceding each tone; for RR, the mean duration of the two cycles just preceding each tone; and for MP, the amplitude in microvolts of the highest tracing occurring in the 5-sec. period preceding each tone. The stimulus level for SR was the lowest level in ohms during a 20-sec. period following the onset of each tone (maximum response level); for HR, the shortest interval between any two consecutive R waves during each tone; for RR, the mean duration of the two cycles immediately following the onset of each tone; and for MP, the amplitude of the highest tracing during each tone. The RR was easy to score, since the tones were given at the peak of inspiration. The R-R interval in HR is displayed as beats per minute by the Fels cardiotachometer.

The amount of background activity (BA) in each of the several systems was determined as follows. The SR and HR records of each *S* in the initial rest period were divided into intervals of 15 sec. For SR, a background interval (BI) was scored if a discrete and transient decrease of at least 800 ohms occurred in the interval. For HR, each 15-sec. interval was divided into 5-sec. subintervals, and a BI was scored if the maximal rate in a 5-sec. subinterval was at least 6 beats/min higher than the maximal rate in the preceding period of 5 sec. The methods used here to score BA in SR and HR are similar to those devised originally by Lacey and Lacey (1958a). The evaluation of BA in MP was comparable to that in HR: A BI was scored if the highest tracing in a given subinterval of 5 sec. was at least twice the highest tracing occurring in the immediately preceding 5-sec. subinterval.

The scoring of BIs for RR was devised in earlier work (Wilson & Dykman, 1960). For each *S* the mean duration and amplitude of all cycles in each 1-min. segment of the initial rest period were derived. Then, any cycle having an inspiratory peak of less than 50% of the mean amplitude for the specific 1-min. interval in which it occurred was incorporated into the previous respiratory cycle. This increased the length of any cycle followed by low-amplitude activity. Finally, any cycle (reconstructed or otherwise) was scored as BA if it deviated in duration at least ±50% from the average cycle of the 1-min. period in which it occurred.

Scaling of autonomic responses.—"Norms" developed in an earlier study (Dykman et al., 1963) were employed. In brief, graphs were constructed in which the stimulus levels of greatest magnitude in each system were plotted for successive prestimulus subranges in experiments involving tones, questions, and digit-recall exercises. Once the highest stimulus levels (lowest value in ohms for SR, shortest cycle time in RR, and fastest rate in HR) were obtained for each of the prestimulus subranges, a line was drawn by inspection to just bound all responses. The lower limit of reactivity was defined arbitrarily, assuming a zero level of response as indicated by no change from the prestimulus level, or by an increase in ohms in SR and deceleration in HR and RR. After these two limiting lines were constructed, the area between them was divided into 10 equal parts with a score of 10 indicating a maximal response. The SR and RR graphs were triangular in shape, fanning out with increasing prestimulus levels. In HR, there was no attenuation of the range of responsiveness at the different prestimulus levels. A comparable scaling method for MP has not yet been developed and in this study the conventional stimulus-prestimulus difference scores were used.

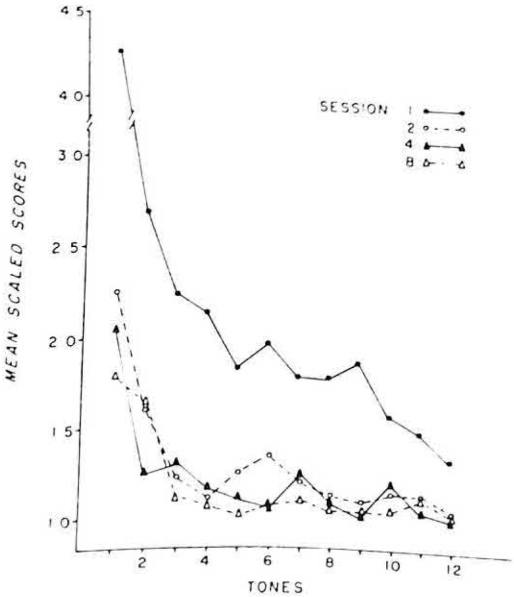


FIG. 1. Mean scale scored in SR (left side) for each tone in each of four sessions.

The graphic method considers the regression of stimulus levels on prestimulus levels and allows for homoscedasticity (Lacey, 1956). As now derived, only significant responses obtain a score of at least two. Of the 1920 responses scored in each system, only 18% of the GSRs, 31% of the HR changes, 30% of the RR changes received a scaled score of two or more. For purposes of contrast, we shall also present the SR, HR, and RR data in terms of *difference scores* (stimulus levels minus prestimulus levels) and *ratio scores* (stimulus levels/prestimulus levels).

RESULTS

Adaptation and extinction.—For convenience in exposition, we arbitrarily defined adaptation as a response decrement within single sessions and extinction as a response decrement between sessions. The use of the term adaptation is questionable. There is, however, some precedent; Lindsley (1951, p. 497) uses it to denote intrasession gradients of the GSR and EEG. On this point it may be worthwhile to point out that all terms commonly used to refer to response diminutions, adaptation, habituation, and extinction, are inadequate in one way or another. Adaptation generally denotes complex adjustments, habituation has as its root habit, and extinction conveys the meaning of the complete disappearance of a response which is rarely (if ever) the case.

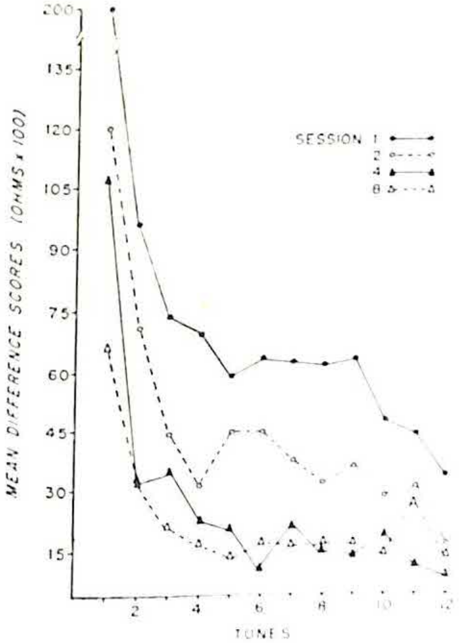


Fig. 2. Mean difference score in SR (left side) for each tone in each of four sessions.

Figure 1 presents the mean scaled score, Fig. 2 the mean difference score, and Fig. 3 the median stimulus prestimulus ratio in SR for each tone in Sessions 1, 2, 4, and 8. Only the left-side curves are shown, but the right-side curves were almost identical. The dotted lines in Fig. 3 represent the function

$$(S/PS)' = 1 - \left[\frac{k}{\left(D^{(0.8)}\right)\left(T^{(0.8)}\right)} \right]$$

where:

$(S/PS)'$ is the predicted value of the median on any trial on any day, D is the session number (1–8), T is the trial number (1–12), k is a constant equal to $1 - (S_1/PS_1)$, and S_1 is the stimulus level and PS_1 the prestimulus level for the first tone on the first day. A number of mathematical equations were applied to the empirical data, but the hyperbolic function appeared by inspection to provide the best fit.

It may be seen that Fig. 1–3 inclusive provide roughly comparable information, although significant events are masked to some extent by the cruder scaled scores.

In order to determine the number of *Ss* who adapted and the number who extinguished in each system in terms of difference scores, scaled scores, and ratio scores, the following procedures were employed. For a given system and kind of score, the mean response to each tone over all sessions was derived for each *S*. The *S* was then considered to have adapted if at least five of the following six criteria were met: the mean response to Tone 1 > the mean response to Tone 10, to Tone 11, and to Tone 12; the mean response to Tone 2 > that to Tone 10, to Tone 11, and to Tone 12. "Reverse" adaptation effects were indicated by smaller mean responses to Tones 1 and 2 than to Tones 10, 11, and 12. Extinction was said to be present if *S* satisfied at least five of the following six criteria: the mean response to all tones on Session 1 > the mean response to all tones on Session 6, on Session 7, and on Session 8; the mean response to all tones on Session 2 > that to all tones on Session 6, on Session 7, and on Session 8. Reverse habituation effects were

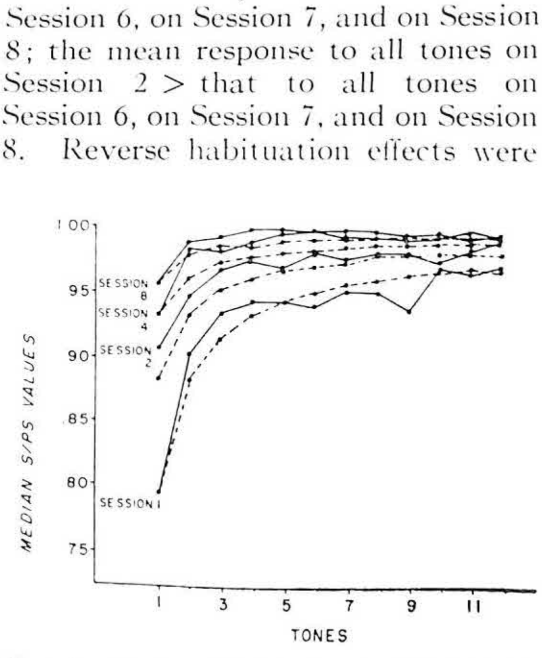


Fig. 3. Median *S/PS* values in SR (left side) for each tone in each of four sessions.

TABLE 1

Number of *S*s Classified According to Conjoint Presence or Absence of Adaptation and Extinction Effects

Classification	System			
	SR	HR	RR	MP
A+E+	13*(13*)	0(0)	2*(2*)	10*(7*)
A+E−	1 (1)	1(1)	0 (0)	1 (1)
A+Eo	5*(5*)	3(4)	2 (2)	8*(5*)
A−E+	0 (0)	0(1)	2*(3*)	0 (0)
A−E−	0 (0)	1(1)	0 (0)	0 (0)
A−Eo	0 (0)	2(2)	2 (2)	0 (0)
AoE+	1 (1)	2(1)	5*(4)	1 (4)
AoE−	0 (0)	4(4)	2 (2)	0 (0)
AoEo	0 (0)	7(6)	5 (5)	0 (3)

* $p < .05$ by binomial expansion.

indicated by smaller mean responses in Sessions 1 and 2 than in Sessions 6, 7, and 8. These criteria, admittedly crude, were designed to be sensitive to a linear trend.

The conjoint presence or absence of adaptation and extinction effects for each *S* in each system was determined as follows. A given *S* was first classified as showing adaptation, reverse adaptation, or no intrasession gradient; and then as showing extinction, reverse extinction, or no intersession gradient. Table 1 shows, for each system, the number of *S*s in each of nine possible classifications. Since the three kinds of response scores provided similar results, the findings presented in Table 1 are based upon difference scores only. Adaptation is denoted by the symbol A+, reverse adaptation by A−, and no gradient by Ao. The E symbols have comparable meanings except that they refer to extinction. It may be seen, e.g., that 13 *S*s adapted and extinguished in SR. Each number in parentheses in Table 1 is a classification of *S*s using Tones 6, 7, and 8 rather than Tones 10, 11, and 12 as the end points in the assessment of adaptation.

It will be noted that in SR and in MP, a significant number of *S*s exhibited both adaptation and extinction, although the conjoint presence of both processes was more pronounced in SR. On the other hand, both SR and MP findings provided some evidence of adaptation in the absence of extinction.

Table 2 presents scaled-score results. The expected percentage of responses with a scaled score of 2.0 or more for each day or each tone is 18% for SR, 31% for HR, 30% for RR. It will be noted that SR adaptation and extinction gradients are similar through the first eight tones and 8 days, and that there is no evidence of systematic adaptation and extinction in HR or RR.

Table 3 shows, of the number of *S*s who reacted more to Tone 1 than to Tone 2 each day in SR, the number who also reacted more to Tone 2 than to Tone 3, to Tone 3 than to Tone 4, etc. The theoretical expectancy for any entry in the first row is 10, while the theoretical expectancy of all other entries is half that of the first row. Considering only the first two tones, there is a significant adaptation

TABLE 2

Percentage of Scaled Scores of 2.0 or Greater by Tones and Days

Tone or Day	SR		HR		RR	
	Tone	Day	Tone	Day	Tone	Day
1	45	50	39	25	34	33
2	28	16	26	30	32	33
3	21	21	28	39	32	31
4	16	13	29	33	28	30
5	16	12	29	29	35	38
6	15	16	31	30	39	24
7	14	7	26	29	30	30
8	11	8	34		29	34
9	15		34		29	
10	10		29		29	
11	11		28		24	
12	13		35		20	

TABLE 3

NUMBER OF *S*s WHOSE RESPONSES IN SKIN RESISTANCE WERE GREATER FOR TONE 1 THAN TONE 2 ON EACH DAY AND ALSO GREATER FOR THE OTHER LISTED TONES ON THE SAME DAY: *S/PS* RATIOS FOR LEFT SIDE OF BODY

Tones Compared	Sessions							
	1	2	3	4	5	6	7	8
1 > 2	19	17	18	19	13	17	16	16
2 > 3	15	10	12	7	7	8	8[a]	9
3 > 4	12	14	11	10	9	11	9[a]	8
3 > 8	13	12	12	11	9	8	7[a]	7
3 > 12	17	15	13	14	7	12	9[b]	7

Note.—$N = 20$ for each cell, unless otherwise indicated.
[a] $N = 19$.
[b] $N = 18$.

gradient on each day for the majority of *S*s.

Figure 4 shows, for the left and right arm separately, the mean difference scores in MP (*S-PS*) for each successive tone (Sessions × *S*s). The mean difference scores for each session appear in Fig. 5. The intersession gradient approaches the hyperbolic form, while the intrasession gradient is essentially linear. The major difference between the adaptation and extinction curves occurs with respect to the first two plotted points. Comparable plots for SR showed the extinction gradient to be steeper than the adaptation gradient, but both curves were clearly hyperbolic in form. Curves constructed for HR and RR failed to provide any evidence of systematic adaptation or extinction.

Mean adaptation and extinction curves in MP were also constructed in terms of log microvolts. For each *S* the log of the mean difference scores for each tone (across sessions) was found, and the mean log for each tone (across *S*s) was then derived. Similarly, for each *S* the log of the mean difference score for each session was found, and the mean log for each session was obtained. The intra- and intersession gradients represented by log microvolts had similar shallow slopes and were linear. Comparable log curves for SR were similar in slope and form to those based upon difference scores.

The possible effect of length of intersession interval (2 or 3 days) upon reactivity in SR and MP was investigated. The factor of

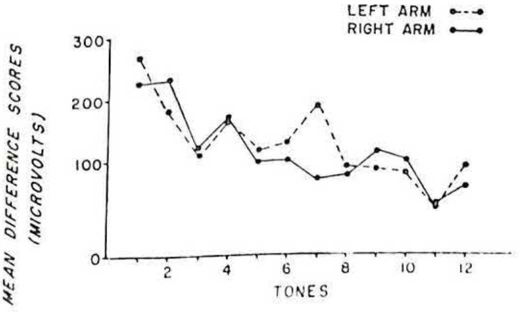


FIG. 4. Mean difference scores in MP for each tone (*S*s × Sessions).

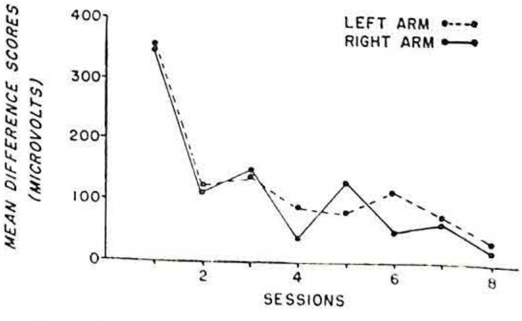


FIG. 5. Mean difference scores in MP for each session (*S*s × Tones).

session number was controlled, since within the 2-day and within the 3-day interval all eight sessions received essentially equal representation. In each system differences in sizes of response to Tone 1 only between the 2- and 3-day intervals were found to be nonsignificant when evaluated by the Mann-Whitney *U* test (Siegel, 1956).

We also attempted to determine whether the reactions in MP were predominantly ORs or some combination of ORs and startle reflexes. Note was made of the number of MP reactions in Session 1 which occurred in .3 sec. or less. Of the 240 responses, only 3.8% had MP activity indicating a startle component.

Intersession decrement in BA. Corah and Stern (1963) found a decrement in BA from one session to a second. Their findings based on a longer test-retest interval than that used by Lacey and Lacey (1958a) or by Wilson and Dykman (1960) were not in accord with predictions from these latter studies.

For a given *S*, session, and system in the present study, there were 20 intervals in which BA could have occurred in the prestimulation rest period. The mean percentage of BIs in each system for each session is shown in Fig. 6. No system exhibited an intersession decrement in BA. The differences in the mean percentage of BIs between Sessions 1 and 2 and Sessions 7 and 8 for each *S* were not reliable ($p > .05$) by Wilcoxon's sign-rank test (Siegel, 1956).

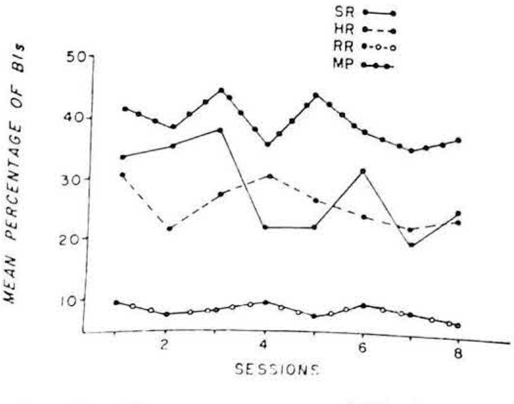


FIG. 6. Mean percentage of BIs in each system and in each session.

TABLE 4

COEFFICIENTS OF CONCORDANCE ACROSS DAILY SESSIONS

System	Stimulus Levels	Scaled Scores	Diff. Scores	*S/PS* Ratios
HR	.62	.47	.48	.52
RR	.70	.59	.51	.52
SR (left)	.67	.33	.44	.42
MP (left)	.53	–	.38	.33

Intersession reliability and intersystem correlations.—We did not compute correlations between levels (stimulus or prestimulus) or between responses in the different systems. Our past work (Dykman et al., 1959, 1963) and work by Lacey (Lacey, 1956; Lacey & Lacey, 1958b) has shown that intersystem correlations reflecting interindividual positioning tend to zero. However, as in the studies just cited, coefficients of concordance (Edwards, 1955) were derived for various measures in each system, and these results appear in Table 4. The *S*s maintained at a reliable level ($p < .001$ in every case) their rank-order position in the group across sessions, regardless of whether adaptation and extinction were present as in SR or MP or absent as in HR or RR. The generally higher coefficients for stimulus levels than for response measures suggests that *S*s seek a constant stimulus level whatever the prestimulus level (Block & Bridger, 1962; Lacey & Lacey, 1962). Coefficients of concordance show that *S*s also maintained at a statistically reliable, but not predictable, level their group positions in BA from one session to the next: .41 for SR, .33 for HR, .36 for RR, and .24 for MP ($p < .01$ for all values).

Table 5 presents intersystem correlations for BA. The basic measure in each system was the mean percentage of BIs across all sessions for each *S*. These correlations suggest two clusters, one consisting of MP, HR, and RR, and the other MP and SR.

Intercorrelations of BIs were also derived for each single session. Reliable positive relationships were found between SR and MP in 3 of the 8 days with low-positive correlations in the remaining sessions; between HR and RR on Day 1 only with low-positive correlations in all other sessions; between SR and RR on Day 5; and between RR and MP on Days 4 and 7 with low-positive relationships in other sessions. These correlations do not reflect a coupling in time: atypical RR cycles rarely coincided with atypical changes in HR (Lacey & Lacey, 1958a; Wilson & Dykman, 1960).

Amount of BA was also studied in relation to response magnitude. Scaled scores in HR correlated .60 ($p < .01$) with BIs in HR. The comparable rank-order correlations for RR and SR were .54 ($p < .02$) and $-.19$ ($p > .10$), respectively. Difference scores in HR correlated .42 ($p < .10$) with BIs in HR. The parallel correlations for RR, SR, and MP were .35 ($p > .10$), .07 ($p > .10$), .45 ($p < .05$), respectively.

TABLE 5

INTERSYSTEM CORRELATIONS OF RESTING BIs

System	SR	HR	RR
HR	−.09		
RR	−.09	.55*	
MP	.57**	.48*	.35

* $p < .05$.
** $p < .01$.

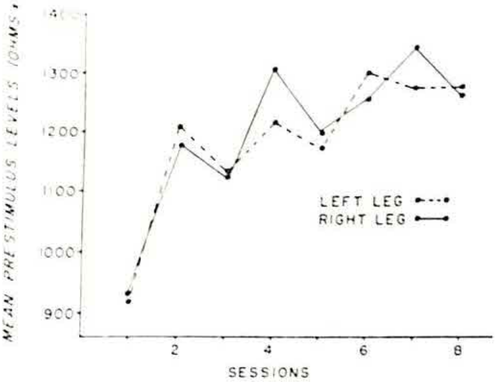


FIG. 7. Mean prestimulus levels of SR in each session (*Ss* × Tones).

Effects of arousal on the OR.—Prestimulus levels were used as an indicator of the background arousal state. Figure 7 is a plot of the mean prestimulus levels in SR (*Ss* × Tones) for each session. The gradually increasing ohms level suggests a general trend for arousal to decrease from day to day. This average trend was representative of most *Ss*.

In HR, the mean prestimulus level decreased from 76 beats/min in Session 1 to 71 beats/min in Session 6, and then increased to 75 beats/min in Session 8. In RR, the mean prestimulus average decreased from 3.96 sec. per cycle to 3.72 sec. per cycle; i.e., rate of breathing accelerated slightly over sessions. This tendency of RR to accelerate has been noted in humans (Davis et al., 1955) and in dogs (Dykman, Mack, & Ackerman, 1965). In MP, there was no gradient; i.e., mean prestimulus levels were approximately equal on all days, either in raw-microvolt or log-microvolt units. The graphs for stimulus levels in SR, HR, and RR were quite similar in form to those obtained for prestimulus levels.

For each system, the mean prestimulus levels to Tones 11 and 12 on a given day were compared to the prestimulus level just preceding Tone

1 on the following day (Wilcoxon's sign-rank test). The tests contrasted levels for Days 1 and 2, Days 2 and 3, and Days 7 and 8. There were no significant findings in HR and RR. Of the three tests run for SR, only the Day 1-Day 2 comparison was significant ($p < .02$); ohms level was higher during the pretone period on Day 2 than during the mean pretone period on Day 1. One significant difference appeared in MP contrast; the pretone level to Tone 1 on Day 8 was higher than the mean pretone level to Tones 11 and 12 on the preceding day ($p < .02$). The evidence is thus negative to the proposition that the generally larger reactions in SR and MP on the first tone each day are related to arousal or drive, as here defined.

We also had scores for each *S* on the Taylor Manifest Anxiety scale (Taylor, 1956). The five *S*s with the highest scores (13–19) gave smaller mean-SR responses on the average than the five *S*s with the lowest anxiety scores (2–7). The curves for the two groups of *S*s were similar in form plotting either median difference scores for each session or median difference scores for each successive trial across all sessions (see Dykman et al., 1959, for corroborating information).

Bilateral differences.—For each *S* in each session bilateral differences in prestimulus levels of SR were evaluated by Wilcoxon's sign-rank test (Siegel, 1956). Six *S*s exhibited reliable differences ($p < .05$) in all sessions, 10 in seven sessions, all 20 in at least five sessions. Across all sessions for all *S*s ($N = 160$), SR was significantly higher on the left in 63 instances and on the right in 78 instances, leaving only 19 sessions in which there was no difference. Six *S*s were variable as to the side showing higher levels, but eight had higher levels on the right and six higher levels on the left in at least five sessions.

Comparable analyses for SR difference scores indicated that only three *S*s had significant left-right differences in at least five sessions. Of the 63 sessions in which left-side levels were significantly higher, 20 sessions showed a significant left-side dominance in response and 2 sessions a right-side dominance. Of the 78 sessions in which right-side levels were significantly higher, 20 sessions showed a significant right-side dominance in response and 7 sessions a left-side dominance. Though there was some concordance in this cross classification of levels and responses, there was more often an absence of response dominance in the presence of reliably higher prestimulus levels, i.e., 41 of 63 cases on the left side and 51 of 78 cases on the right. Neither differences in response nor levels were related to handedness. In Session 1 the median absolute difference between prestimulus levels on the two sides of the body was 4,380 ohms, with the middle 50% of the observations ranging from 3,000 to 11,280 ohms. The median difference and the middle 50% range showed a general increase over subsequent sessions. By Session 8 the median difference was 11,550 ohms, with a range of 5,880 to 23,540 ohms for the middle 50% of the observations.

Comparable analyses were carried out for MP. Across all 160 sessions there were 47 in which prestimulus levels on the left were reliably higher than those on the right, 25 sessions in which the opposite was true, and 88 sessions in which there were no significant differences in levels. The major feature here was that of similarity of prestimulus levels on the two sides of the body; only three *S*s exhibited higher levels on the same side of the body in at least five sessions. The findings for difference scores in MP were in essential agreement with those for difference scores in SR in that no *S* evidenced consistent bilateral differences over sessions. Furthermore, size of response was independent of the prestimulus level in MP.

Discussion

This study was designed to test the hypothesis that adaptation and extinction are identical processes; i.e., in this orienting situation, one term—whether adaptation, habituation, or extinction—should be used to describe both intra- and intersession decreases in reactivity. It was assumed that this hypothesis would be supported if the mean-trial gradient was similar to the mean-session gradient in each of the several functions studied. And, as a more stringent test, it was expected that *Ss* who adapted in a given system would extinguish in that system and that *Ss* who failed to adapt would not extinguish. Knowing the partial dependence of biological responses on uncontrolled internal factors, strict conformity to the hypothesis was not expected.

The reader may reason that the hypothesis is invalid since adaptation and extinction are defined alike except for variation in the time of presentation of stimuli. But this variation in time could be important in discriminating true extinction from response decrements attributable to drive reduction. Our supposition, right or wrong, was that the intrasession effect might possibly depend more on drive reduction than the intersession effect and that the latter is the better criterion of true extinction. If it were then shown that these two processes as defined are similar and that drive was not a critical factor, one should use one term to describe both functions.

Our results are interpreted as supporting, in general, the concept that adaptation and extinction are equivalent processes, and particularly in SR, since the exponential constant for trials was equal to the exponential constant for days. Skin resistance, the system in which *Ss* were most reactive, provided the best evidence of systematic adaptation and extinction followed in descending order by MP, RR, and HR.

The recovery of SR and MP reactions from one day to the next might be explained in several not incompatible ways (*a*) as the dissipation of conditioned inhibition through disuse; (*b*) as a warm-up effect in which a trial or two are required to sensitize inhibitory mechanisms; and (*c*) as a procedure of reinforcement in which the first trial each day reassures *Ss* of the continuing absence of noxious stimulation (Dykman, Gantt, & Whitehorn, 1956). Although we could not evaluate these different possibilities, the latter is somewhat suspect from the standpoint of our data on generalized drive (Hull, 1951) and anxiety (Spence, 1956). The generally larger responses in SR and MP on the first trial each day were not related to prestimulus levels, one measure of generalized drive. Anxiety as measured by the Taylor scale was not related to intra- and intersession decrements. And *Ss* with the highest scores on the anxiety test actually gave the smallest GSRs (cf. Dykman et al., 1959).

Some recent unpublished research in our laboratory indicates that background activity in SR decreases as *Ss* are going to sleep and increases when they are awakened. Thus, background activity does vary with significant changes in the level of drive (or arousal). By contrast, the absence of habituation in background activity from day to day in the present work indicates that the drive level was relatively constant. Thus, here again, there is evidence that drive or arousal did not enter as a critical variable in shaping response diminutions in the present experiment.

The formula derived for SR was for the *left* side of the body only. Mean graphs in SR and MP constructed on the basis of stimulus-prestimulus differences were, however, highly similar for the two sides of the body. This similarity was not impaired by rather sizable differences in the prestimulus levels on the two sides. Thus responses in this situation were relatively independent of prestimulus level, despite the predictions of the law of initial values (Lacey, 1956). The appreciable bilateral differences in normal *Ss* were surprising. (Our reason for recording bilaterally was to obtain some normative data to contrast with patient data in other studies.)

Turning to broader considerations, the OR is seen as a gross-organismic reaction including multiple skeletal-motor and autonomic components and depending very much on the individual differences in the innate organization of responses. The OR is an appropriate reaction to an attention-getting stimulus, from inside or outside the organism, which is unusual in nature or unexpected in time. This definition may appear not to encompass the reactions to tones in the present experiment. However, pure tones are unusual stimuli and as they recur provide a sharp contrast to the background conditions of quiescence. Furthermore, the exact onset of tones could not be anticipated by *Ss*. Startle components, while generally absent here, may precede the OR depending upon the set of *S*, his level of arousal, and the intensity of the stimulus, variables which also underlie the vigor of the OR. The possible interference with or augmentation of the OR by the startle reflex deserves detailed investigation.

The specific stimulus-directed movements of the OR are more readily inhibited than the various autonomic and motor components. The OR appears to undergo a faster extinction in the human than in the dog (Dykman et al., 1959). Its unusual persistence in certain primates is perhaps attributable to differences in degree of domestication. Wild rabbits have more pronounced ORs than pets (personal observation). Current evidence indicates that the OR is innate; it is present in the neonate a few hours after birth (Wertheimer, 1962). Further, its nature and resistance to extinction vary with phylogeny (cf. Razran, 1961, p. 114).

A number of writers have suggested the probable importance of the OR in conditioning (Grings, 1960, 1963; Sokolov, 1960; Stern, Stewart, & Winokur, 1961; Stewart, Stern, Winokur, & Fredman, 1961). It is known from earlier work (Dykman et al., 1956) that the first UCS reinstates and augments the general bodily components of the OR but not generally the specific motor orientation to signals. The early positive CR appears to be an amalgamation of the OR to the CS and certain nonspecific UCS reactions (e.g., looking down at shock electrodes, restlessness, looking from side to side). As the CR develops, most, but not all, of the general orienting components to the CS and the UCS and certain other nonspecific UCS reactions undergo extinction. The general bodily components of the OR may serve as important mediators of later more specific responses. The theoretical model implied by these later statements, to be elaborated in greater detail elsewhere by Dykman, accords with the general psychological notion that the early phase of conditioning is one of drive induction or emotional sensitization.

REFERENCES

Allen, C. K., Hill, F. A., & Wickens, D. D. The orienting reflex as a function of the interstimulus interval of compound stimuli. *J. exp. Psychol.*, 1963, **65**, 309–316.

Bitterman, M. E., & Holtzman, W. H. Conditioning and extinction of the GSR as a function of anxiety. *J. abnorm. soc. Psychol.*, 1952, **47**, 615–623.

Block, J. D., & Bridger, W. H. The law of initial value in psychophysiology: A reformulation in terms of experimental and theoretical considerations. *Ann. N. Y. Acad. Sci.*, 1962, **98**, 1229–1240.

Corah, N. L., & Stern, J. A. Stability and adaptation of some measures of electrodermal activity in children. *J. exp. Psychol.*, 1963, **65**, 80–85.

Darrow, C. W. Differences in the physiological reactions to sensory and ideational stimuli. *Psychol. Bull.*, 1929, **26**, 185–201.

Davis, R. C. Response and adaptation to brief noises of high intensity. *USN SAM Res. Rep.*, 1953, Contract AF 33(038)-19630.

Davis, R. C., Buchwald, A. M., & Frankmann, R. W. Autonomic and muscular responses, and their relation to simple stimuli. *Psychol. Monogr.*, 1955, **69**(20, Whole No. 405).

Dykman, R. A., Ackerman, P. T., Galbrecht, C. R., & Reese, W. G. Physiological reactivity to different stressors and methods of evaluation. *Psychosom. Med.*, 1963, **25**, 37–59.

Dykman, R. A., & Gantt, W. H. The parasympathetic component of unlearned and acquired cardiac responses. *J. comp. physiol. Psychol.*, 1959, **52**, 163–167.

Dykman, R. A., Gantt, W. H., & Whitehorn, J. C. Conditioning as emotional sensitization and differentiation. *Psychol. Monogr.*, 1956, **70**(15, Whole No. 422).

Dykman, R. A., Mack, R. L., & Ackerman, P. T. The evaluation of autonomic and motor components of the nonavoidance conditioned response in the dog. *Psychophysiology*, 1965, **1**, 209–230.

Dykman, R. A., Reese, W. G., Galbrecht, C. R., & Thomasson, P. J. Psychophysiological reactions to novel stimuli: Measurement, adaptation, and relationship of psychological and physiological variables in the normal human. *Ann. N. Y. Acad. Sci.*, 1959, **79**, 43–107.

Edwards, A. L. *Statistical methods for the behavioral sciences.* New York: Rinehart, 1955.

Grings, W. W. Preparatory set variables related to classical conditioning of autonomic responses. *Psychol. Rev.*, 1960, **67**, 243–252.

Grings, W. W. Classical conditioning. In M. H. Marx (Ed.), *Theories in contemporary psychology.* New York: Macmillan, 1963. Pp. 495–525.

Howe, E. S. GSR conditioning in anxiety states, normals, and chronic functional schizophrenic patients. *J. abnorm. soc. Psychol.*, 1958, **56**, 183–189.

Hull, C. L. *Essentials of behavior.* New Haven: Yale Univer. Press, 1951.

Johnson, L. C. Some attributes of spontaneous autonomic activity. *J. comp. physiol. Psychol.*, 1963, **56**, 415–422.

Lacey, J. I. The evaluation of autonomic responses: Toward a general solution. *Ann. N. Y. Acad. Sci.*, 1956, **67**, 123–164.

Lacey, J. I., & Lacey, B. C. The relationship of resting autonomic activity to motor impulsivity. In, *The brain and human behavior.* Baltimore: Williams & Wilkins, 1958. Pp. 144–209. (a)

Lacey, J. I., & Lacey, B. C. Verification and extension of the principle of autonomic response stereotypy. *Amer. J. Psychol.*, 1958, **71**, 50–73. (b)

Lacey, J. I., & Lacey, B. C. The law of initial value in the longitudinal study of autonomic constitution: Reproducibility of autonomic responses and response patterns over a four-year interval. *Ann. N. Y. Acad. Sci.*, 1962, **98**, 1257–1290, 1322–1326.

Lindsley, D. B. Emotion. In H. S. Langfeld (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 473–516.

Pavlov, I. P. *Conditioned reflexes.* (Trans. & ed. by G. V. Anrep) Oxford: Oxford Univer. Press, 1927.

Rachman, S. Reliability of galvanic skin response measures. *Psychol. Rep.*, 1960, **6**, 326.

Razran, G. The observable unconscious and the inferable conscious in current Soviet psychophysiology: Interoceptive conditioning, semantic conditioning, and the orienting reflex. *Psychol. Rev.*, 1961, **68**, 81–147.

Robinson, J., & Gantt, W. H. The orienting reflex (questioning reaction): Cardiac, respiratory, salivary, and motor components. *Bull. Johns Hopkins Hosp.*, 1947, **80**, 231–253.

Siegel, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

Sokolov, E. N. Neuronal models and the orienting reflex. In M. A. B. Brazier (Ed.), *Central nervous system and behavior.* New York: Josiah Macy, Jr. Found., 1960. Pp. 187–276.

Sokolov, E. N. *Perception and the conditional reflex.* Oxford: Pergamon, 1963.

Spence, K. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

Stern, J. A., Stewart, M. A., & Winokur, G. An investigation of some relationships between various measures of galvanic skin response. *J. Psychosom. Res.*, 1961, **5**, 215–223.

Stewart, M. A., Stern, J. A., Winokur, G., & Fredman, S. An analysis of GSR conditioning. *Psychol. Rev.*, 1961, **68**, 60–67.

Taylor, J. A. Drive theory and manifest anxiety. *Psychol. Bull.*, 1956, **53**, 303–320.

Unger, S. M. Habituation of the vasoconstrictive orienting reaction. *J. exp. Psychol.*, 1964, **67**, 11–18.

Wertheimer, M. Psychomotor coordination of auditory and visual space at birth. *Science*, 1962, **134**, 1692.

Wilson, J. W. D., & Dykman, R. A. Background autonomic activity in medical students. *J. comp. physiol. Psychol.*, 1960, **53**, 405–411.

(Received July 15, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 6, 598–601

EFFECTS OF THE VISUAL FIELD UPON PERCEPTION OF CHANGE IN SPATIAL ORIENTATION[1]

NORMAN L. CORAH[2]

Washington University, St. Louis

The study investigated the effects of a frame and its orientation upon the perceived change in orientation of a rod target rotated at a speed below threshold for movement. 84 Ss were equally divided among 7 conditions. The results demonstrated that field structure was generally less important than starting position of the target. Change was more readily perceived when initial target position was at the vertical than when it was tilted ($p < .05$). A tilted field with the target rotating away from the vertical produced the greatest lag in detection. The results are interpreted in terms of adaptation-level theory.

Many studies in recent years have investigated the effects of such variables as posture and visual field structure on judgments of spatial orientation—especially judgments of the perceived vertical. However, there is a relative paucity of data on the perception of change in spatial orientation in the frontal parallel plane.

Some of the human engineering studies of instrument-dial arrangements may be considered relevant here. Aircraft instrument dials as well as others usually have the starting points of their scales in one of the four "cardinal" positions. There appears to be no superiority of any one of these positions in the check reading of instrument groups (White, Warrick, & Grether, 1953). Moreover, shifting the starting point of a scale to other than a cardinal position does not appear to affect the accuracy with which the instruments can be checked (Chapanis, Garner, & Morgan, 1949). However, when there are relatively few visual reference lines present, somewhat different results are obtained. Judgments of the angle formed by a line and a pointer are most accurate at the 0° and 90° points. The 45° point appears to be next in accuracy of reading (Fitts, 1951).

The most relevant study in this area is one by Tepas (1960). He investigated the effects of visual-field structure, target structure, starting position, and instructional set on the perceived change in orientation of a target rotated at a speed which was too slow for movement to be perceived. Vertical and 25° tilted starting positions were used. Field structure was varied by presenting the target in a Ganzfeld or in a lighted room. Perceived change in orientation of the target was best when a vertical starting position was used. Visual field structure did not appear to influence judgments of change except through interaction with instructional set.

The present study was concerned with the effects of target starting position and visual field structure on perceived change in orientation.

---

[1] An earlier version of this paper was read at the Midwestern Psychological Association meeting in St. Louis, May 1, 1964. The writer acknowledges the assistance of Joan Clinch in the collection of the data. The data analysis was facilitated by support of the Washington University Computer Facilities through National Science Foundation Grant G-22296.

[2] Now at the School of Dentistry, State University of New York at Buffalo.

METHOD

Subjects.—The *Ss* were 42 men and 42 women from general introductory psychology classes at Washington University. The *Ss* were randomly assigned to one of seven groups with the restriction that six men and six women comprise each group. There were, consequently, seven groups of 12 *Ss* each.

Apparatus.—The standard Witkin rod and frame were used (Witkin, Lewis, Hertzman, Machover, Meissner, & Wapner, 1954). The frame could be rotated in the usual fashion by hand. The rod was rotated by an electronic motor control unit. This unit consisted of a 1/50 hp. dc shunt wound motor having a gear shaft loaded with an 1120:1 ratio and operated by a Heller 2T60 Thyratron tube motor controller unit. Speed was further reduced by connecting a 25:1 Boston gear box between the motor and the shaft of the rod. For all conditions, the rod was rotated at a relatively constant speed of 4.9° per minute. This speed is below the absolute threshold for the perception of movement. Noise from the motor was masked by a fan.

Procedure.—The *S* was blindfolded and brought into the laboratory where he was seated upright with his head positioned in a clamp 7 ft. from the rod and frame. The apparatus was described and *S* was told that he should observe the rod and report any change in orientation or tilt, if it should occur, and its direction. It was also indicated that several minutes might pass before any change might occur. At the beginning of each trial, *S* removed the blindfold and observed the rod. When he reported a change, the apparatus was stopped and *S* replaced the blindfold. Four trials were administered to each *S*. The mean degrees of rotation before a change was detected was the measure used.

The presence or absence of the frame and its tilt provided the conditions of visual field structure. The starting positions in combination with field structure were as follows: Group 1—Rod vertical, no frame; Group 2—Rod tilted 28°, no frame; Group 3—Rod vertical, frame vertical; Group 4—Rod tilted 28°, frame vertical; Group 5—Rod vertical, frame tilted 28°; Group 6—Rod and frame both tilted 28°; Group 7—Rod tilted 28°, frame tilted 38°.

When the rod started at the vertical, it was rotated away from the vertical. When it started in a tilted position, it was rotated towards the vertical. Two trials to the right and two to the left were given in RLLR or LRRL sequence. Under Cond. 6 and 7 where both rod and frame were tilted, they always began tilted to the same side.

TABLE 1

DEGREES OF ROTATION BEFORE CHANGE WAS REPORTED

Group and Starting Position	Mdn.	*M*	*SD*
1. RV	3.18	3.27	1.28
2. RT	6.12	6.22	3.27
3. RV, FV	3.00	2.96	.81
4. RT, FV	6.12	5.75	2.40
5. RV, FT	8.56	9.50	3.90
6. RT, FT	5.68	5.58	3.46
7. RT, FTT	5.56	5.29	2.53

Note.—RV = rod vertical; RT = rod tilted 28°; FV = frame vertical; FT = frame tilted 28°; FTT = frame tilted 38°.

RESULTS

Table 1 gives the median, mean, and *SD* of degrees of rotation before change was perceived for each group. Hartley's largest *F* ratio yielded an $F_{\max}$ (7, 11) = 23.09, $p < .01$, indicating marked heterogeneity of variance (Walker & Lev, 1953). Consequently, a nonparametric analysis was used. A Kruskal-Wallis one-way analysis of variance (Siegel, 1956) yielded an *H* of 32.27 which is significant beyond the .001 level ($df = 6$).

Mann-Whitney *U* tests (Siegel, 1956) calculated among the groups are presented in Table 2. These results suggest that the most important variable was the starting position

TABLE 2

MANN-WHITNEY *U* TESTS AMONG THE GROUPS

Group	2	3	4	5	6	7
1	26.5*	63.5	27.5*	8.0*	29.5*	32.0*
2		21.5*	70.0	37.0*	58.5	61.5
3			24.0*	7.0*	25.0*	29.5*
4				26.0*	56.0	59.5
5					22.0*	20.5*
6						71.0

* $p < .05$.

TABLE 3

F RATIOS AMONG THE GROUP VARIANCES

Group	2	3	4	5	6	7
1	6.52*	2.48	3.52*	9.29*	7.29*	3.90*
2		16.21*	1.85	1.42	1.12	1.67
3			8.76*	23.09*	18.12*	9.70*
4				2.64	2.07	1.11
5					1.27	2.38
6						1.87

* $p < .05$.

of the rod. Perception of change was significantly worse when the rod was tilted than when it was vertical. The visual framework had a significant effect upon judgments only in Cond. 5 where the rod started at the vertical and the frame was tilted. The performance of this group was significantly worse than that of all other groups. It may also be noted that the presence of the frame in Cond. 3 produced slightly better, although not significantly better, performance than Cond. 1 where it was absent.

Table 3 presents the *F* ratios among the group variances. The variability of judgments is consistently greater for those conditions where the rod started in the tilted position. There was a slight but not significant tendency for greater variability to occur in the conditions where the frame was absent.

DISCUSSION

In the present study, the starting position of the rod is in effect an anchor or comparison stimulus. However, this anchor is an internal representation since the position of the moving rod is being compared to its own starting position. If Cond. 5 is temporarily disregarded, the results indicate that when the rod started at the vertical, discrimination of change was much more effective than when a tilted starting position was used. This result would suggest that the phenomenal vertical plays an important role in maintaining the internal reference or anchor. The concept of a subjective horizontal-vertical coordinate system has been advanced by Onley and Volkmann (1958) in their study of perpendicularity and has also been suggested in the work of Fitts (1951) and Tepas (1960).

Tepas (1960) has suggested that phenomena of perceptual space may be interpreted in terms of Helson's (1959) adaptation-level (AL) theory. In this context, the present results would suggest that *S*'s kinesthetic sensations of the vertical and his past experience with the normal coordinates of perceptual space contribute heavily to the AL when a vertical starting position is used and minimally to the AL with a tilted starting position. Thus, the concern of the present study is with the neutral or indifference range of the AL in which no change is perceived as a function of factors contributing to the stability of the AL.

In the results considered so far, the AL is roughly equivalent to the target starting position. However, Cond. 5 produced an exception to these results. It may be noted that this condition is closest to that used by Witkin (Witkin et al., 1954) in his studies of the effects of the tilted frame on judgments of the vertical. This procedure has the effect of displacing the judged vertical an average of 5° or 6° from the true vertical. Even if the frame in this condition produced a shift in the phenomenal vertical and, consequently, the AL, the neutral range for the anchor should have been narrower than that existing under conditions where the rod was tilted 28°. Yet, detection of change was significantly worse in Cond. 5 than under a tilted starting position.

There are two possibilities which may explain this result in Cond. 5. If the frame produces a shift in the AL towards the frame, the rod may start at the upper limit of the indifference range while it is in the middle of this range (at the AL) under the other conditions. This alternative would suggest that setting the rod at the judged vertical, as determined by the Witkin procedure, would result in discriminations no greater than those obtained under the other rod-vertical conditions.

Another possibility may be that the tilted frame makes the phenomenal vertical unstable and reduces or eliminates its contribution to the AL. The greater variability in judgments associated with Cond. 5 would tend to support this idea. Either one or some combination of these possibilities would account for the results. Only further research directed at the problem can decide the issue.

REFERENCES

CHAPANIS, A., GARNER, W. R., & MORGAN, C. T. *Applied experimental psychology.* New York: Wiley, 1949.

FITTS, P. M. Engineering psychology and equipment design. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 1287–1340.

HELSON, H. Adaptation level theory. In S. Koch (Ed.), *Psychology: A study of a science. Vol. 1. Sensory, perceptual and physiological formulations.* New York: McGraw-Hill, 1959. Pp. 565–621.

ONLEY, J. W., & VOLKMANN, J. The visual perception of perpendicularity. *Amer. J. Psychol.*, 1958, **71**, 504–516.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

TEPAS, D. I. The influence of field articulation and set on spatial orientation. Unpublished doctoral dissertation, University of Buffalo, 1960.

WALKER, H. M., & LEV, J. *Statistical inference.* New York: Holt, 1953.

WHITE, W. J., WARRICK, M. J., & GRETHER, W. F. Instrument reading III: Check reading of instrument groups. *J. appl. Psychol.*, 1953, **37**, 302–307.

WITKIN, H. A., LEWIS, H. B., HERTZMAN, M., MACHOVER, K., MEISSNER, P. B., & WAPNER, S. *Personality through perception.* New York: Harper, 1954.

(Received July 20, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 6, 602–605

DEPENDENCE OF EQUALITY JUDGMENTS UPON THE TEMPORAL INTERVAL BETWEEN STIMULUS PRESENTATIONS[1]

WALLACE R. McALLISTER, DOROTHY E. McALLISTER,
AND JOSEPH J. FRANCHINA[2]

Syracuse University

Judgments of equality were obtained from 8 independent groups of *S*s following a comparison of a standard 50-mm. line with either a 50-, 54-, 58-, or 62-mm. line presented after either ½ min. or 8 min. Percentage of equal judgments was significantly greater following the 8-min. presentation interval than following the ½-min. interval only for lines longer than the standard. A 2nd judgment by each S with the other presentation interval indicated that the effect of presentation interval depended upon the order of intervals used. The 1st-judgment data parallel the results of several animal studies which showed an increase in response strength to generalized stimuli with time. All of these findings are consistent with the hypothesis that the number of jnd's separating 2 similar stimuli decreases as the time between presentations of the stimuli increases. This hypothesis reconciles the data with Hull's theory of stimulus generalization.

One possible method of explaining, within a Hullian framework, the finding that responses to generalized stimuli increase in strength with time (McAllister & McAllister, 1963; Perkins & Weyant, 1958; Saltz & Asdourian, 1963; Thomas & Lopez, 1962) is to assume a decrease with time in the number of jnd's along the stimulus dimension. That is, in the scaling of a given range of physical stimuli, more jnd's would be obtained if a short rather than a long time interval was used between presentations of the standard and comparison stimuli. This hypothesis modifies Hull's statements (1943) about stimulus generalization only by introducing time as a variable in the psychophysical scaling of the stimulus dimension. It, thus, avoids, in accounting for the above results, the necessity of assuming, contrary to theory, that habit strength to generalized stimuli increases spontaneously with time.

All of the above studies employed animal *S*s. To determine the generality of the empirical finding and to manipulate a stimulus dimension systematically, the present study employed a psychophysical procedure with human *S*s. The hypothesis predicts that two similar physical stimuli would be more likely to be judged equal the greater the length of time between their presentations.

METHOD

The *S*s were 117 men and 155 women undergraduate students at Syracuse University. A psychophysical method described by Woodworth and Schlosberg (1954, p. 210) was used. The stimuli were black lines, $\frac{1}{32}$ in. wide and either 50, 54, 58, or 62 mm. long centered singly on white cardboard, 11 × 14 in.

The *S*s were instructed that the experiment was on perceptual discrimination and that they would be shown one line and, at a later time, another line and that their task was to judge whether the second line was longer than, or equal in length to, the first line. At the start of the experiment two cardboards were placed face down in front of *S*. When *E* said "Ready," *S* took hold of the nearest cardboard and at the signal, "Look," raised it and looked at the line. After 10 sec., *E* said, "Stop," and *S* turned the cardboard face down and read a magazine. Either ½ min. or 8 min. later the procedure was repeated for the second cardboard of the pair, after which *S* recorded his judgment of equal or longer. The cardboards and data were then removed, and *S* read a magazine for 10 min. Meanwhile *E* placed two new cardboards before *S*. The judging procedure was then repeated exactly as described above except that the other presentation interval was used.

---

[1] This investigation was supported in part by a research grant, MH-02064, from the National Institute of Mental Health, Public Health Service.

[2] Now at Southern Methodist University.

For each *S*, the first line observed for each of the judgments was 50 mm. long, the standard, and the second, or comparison, line was on both occasions either 54, 58, or 62 mm. Thus, there were eight groups of 34 *S*s each, four with Order 1 (½ min.–8 min.) and four with Order 2 (8 min.–½ min.). Within each order, the groups varied with respect to the lengths, in millimeters, of the lines compared (50–50, 50–54, 50–58, or 50–62).

A maximum of four *S*s participated in the experiment at one time. The two orders of presentation intervals were assigned alternately to successive groups of *S*s. Within a group, *S*s were randomly assigned to judge one of the four sets of line lengths.

These procedures were adopted so as to provide data comparable to that of the animal study from this laboratory (McAllister & McAllister, 1963). Thus, it was required that there be independent *S*s at each generalization point at each temporal interval and that the procedure permit the analysis of data which was not contaminated by a reexposure to the training stimulus. A study by Lipsitt and Engen (1961) bears on the present issue but does not meet these requirements.

RESULTS

Figure 1 presents the first-judgment data from each of the two orders. Thus, each point represents the data from 34 different *S*s. It is these data which are most relevant to the purpose of this paper. As will be shown later, second judgments in this situation are contaminated by transfer effects.

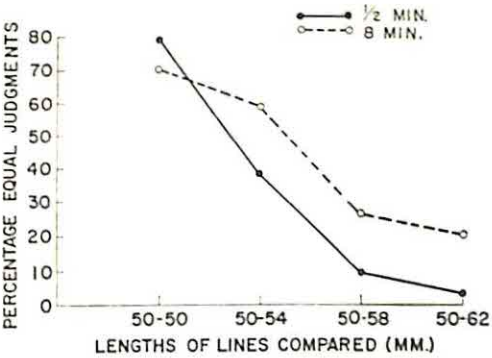


FIG. 1. Percentage of equal judgments as a function of lengths of lines compared for the ½- and 8-min. presentation intervals. (Only the first judgment of each *S* is plotted.)

The decrease in the percentage of equal judgments as the length of the comparison stimulus increased was significant beyond the .001 level both for the ½- and for the 8-min. presentation intervals, χ^2 (3) = 56.98, and 24.58, respectively.

The difference between the presentation intervals for each length of the comparison stimulus was evaluated with χ^2 separately except that the two longest stimuli were combined because of small theoretical frequencies. When the lines were both 50 mm. in length, the difference in equal judgments between the two temporal conditions was not significant, χ^2 (1) = .31, $p > .50$.[3] These results suggest that there is no increase in the tendency for equal judgments to occur merely as a function of time. Therefore, the hypothesized effect of the temporal interval upon the judgments of lines differing in length may be clearly tested. The difference between the 50–54-mm. line groups was not significant, χ^2 (1) = 2.12, $p > .10$. However, for the 50–58 and 50–62 mm. lines combined, the difference was significant beyond the .01 level,

---

[3] Throughout this paper all χ^2 values with 1 *df* were obtained using Yates' correction except that when several χ^2 values were to be combined the uncorrected formula was used since the additive principle does not apply to corrected values (Lewis & Burke, 1949, p. 463).

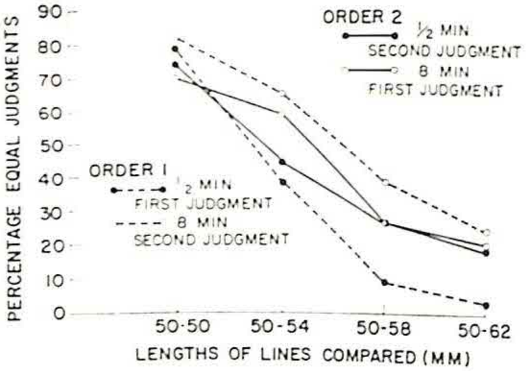


FIG. 2. Percentage of equal judgments as a function of lengths of lines compared for the ½- and 8-min. presentation intervals for Order 1 and Order 2.

χ^2 (1) = 7.09. For both analyses, a greater percentage of equal judgments occurred with the 8-min., than with the ½-min., interval. That the overall increase in equal judgments with an increase in presentation interval is significant is indicated by combining the separate χ^2 values, χ^2 (2) = 11.33, $p < .005$. While the conclusion from these data applies to the lengths of lines employed here, it should be noted that if sufficiently long comparison stimuli were used, the curves should converge at zero. However, concomitant increases in the lengths of the presentation intervals would be expected to affect the point of convergence.

The effect of order, as well as presentation interval, upon judgments of equality is illustrated in Fig. 2. It is apparent that when the stimuli were physically equal (50 mm. compared with 50 mm.) neither variable had a marked effect upon the judgments. Within Order 1, as well as within Order 2, the number of *Ss* giving equal judgments at the ½-min. and at the 8-min. presentation intervals was almost identical (27 vs. 28 and 25 vs. 24, respectively). For neither order was the binomial test (McNemar, 1962, p. 54) significant ($p = 1.00$ in each case). Likewise, the difference between the two orders was not significant at either the ½-min., χ^2 (1) = .08, $p > .70$, or the 8-min., χ^2 (1) = .73, $p > .30$, presentation interval.

When the comparison stimuli were longer than the standard, there was a marked interaction between order and presentation interval. As may be noted in the figure, the effect of presentation interval obtained with Order 1 paralleled the first-judgment data shown in Fig. 1 while that with Order 2 did not. There was a tendency for an increase in the percentage of equal judgments to occur for the second judgments regardless of the order of presentation intervals. Specifically, considering the two ½-min. curves, there was an increase in the percentage of equal judgments when this judgment was second. A similar result may be observed for the 8-min. condition. As a consequence, the difference between the presentation-interval curves was increased for Order 1 and decreased for Order 2. A two-tailed binomial test indicated that for Order 1 the percentage of equal judgments was significantly greater for the 8-min. than for the ½-min. presentation interval for the 50–54, 50–58, and 50–62 comparisons ($p = .01$, .02, and .04, respectively). Comparable tests carried out for Order 2 yielded p values of .42, 1.00, and 1.00. The differences between Order 1 and Order 2 for the 8-min. presentation interval for the 50–54, 50–58, and 50–62 comparisons were not significant, χ^2 (1) = .06, .60, 0; $p > .80$, $> .30$, and 1.00, respectively. Similar analyses for the ½-min. presentation interval yielded a nonsignificant value for the 54-mm. comparison stimulus, χ^2 (1) = .06, $p > .80$, but a significant difference for the 58 and 62-mm. stimulus conditions, combined because of small theoretical frequencies, χ^2 (1) = 6.12, $p < .02$.

DISCUSSION

The first-judgment data indicate that a generalized stimulus is judged to be more similar to a standard stimulus after 8 min. than after ½ min. Similar results were obtained by Lipsitt and Engen (1961) using much shorter presentation intervals and somewhat different procedures. Analogously, the results of the animal studies indicate that a response to a generalized stimulus is more like that to a training stimulus the greater the time elapsing since exposure to the training stimulus. These findings are consistent with the hypothesis that the number of jnd's separating two similar stimuli decreases as the presentation interval increases. Assuming this hypothesis, the notion of a stable generalized habit strength may be maintained provided that the stimulus dimension is specified in psychophysical units which have been obtained using time as a parameter in the scaling procedure. Thus, equal generalization would be expected with different time intervals between training and testing if more dissimilar stimuli were employed as the temporal interval increased. For instance, from Fig. 1, following training to a 50-mm. line, the response to a 53-mm. line after ½ min. and to a 55-mm. line after 8 min. would be expected to be equal. The obtained effect of order suggests that caution must be used if scaling procedures are employed in the proposed manner.

REFERENCES

HULL, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

LEWIS, D., & BURKE, C. J. The use and misuse of the chi-square test. *Psychol. Bull.*, 1949, **46**, 433–489.

LIPSITT, L. P., & ENGEN, T. Effects of presentation of paired and single-stimulus on discrimination of length. *Amer. J. Psychol.*, 1961, **74**, 274–277.

MCALLISTER, W. R., & MCALLISTER, D. E. Increase over time in the stimulus generalization of acquired fear. *J. exp. Psychol.*, 1963, **65**, 576–582.

MCNEMAR, Q. *Psychological statistics.* New York: Wiley, 1962.

PERKINS, C. C., JR., & WEYANT, R. G. The interval between training and test trials as a determiner of the slope of generalization gradients. *J. comp. physiol. Psychol.*, 1958, **51**, 596–600.

SALTZ, E., & ASDOURIAN, D. Incubation of anxiety as a function of cognitive differentiation. *J. exp. Psychol.*, 1963, **66**, 17–22.

THOMAS, D. R., & LOPEZ, L. J. The effects of delayed testing on generalization slope. *J. comp. physiol. Psychol.*, 1962, **55**, 541–544.

WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology.* New York: Holt, 1954.

(Received July 20, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 6, 606–610

ANAGRAM SOLUTION TIMES: A FUNCTION OF INDIVIDUAL DIFFERENCES IN STORED DIGRAM FREQUENCIES

M. E. TRESSELT AND M. S. MAYZNER

Department of Industrial Engineering and Operations Research, New York University

2 groups of 20 Ss each, selected on the basis of their scores (high vs. low) on a skeleton-word test designed to assess individual differences in stored digram frequencies, were presented with a set of 15 anagrams. Anagram letter order, anagram digram-frequency totals, and solution-word Thorndike-Lorge frequencies were held constant, only solution-word, digram-frequency totals varied, i.e., 5 anagrams had low totals, 5 anagrams had medium totals, and 5 anagrams had high digram-frequency totals. The results showed that (a) anagram solution times increased systematically as solution-word, digram-frequency totals decreased and (b) Ss with high-stored digram-frequency scores solved anagrams significantly faster than Ss with low-stored digram-frequency scores. The findings were interpreted as supporting an S-R mediational model of problem-solving behavior.

In several earlier studies (Mayzner & Tresselt, 1958, 1959, 1962a, 1963, 1964) the effects of a number of variables on anagram solution times were examined within the framework of an S-R mediational model of problem-solving behavior (Kendler, 1961; Kendler & Kendler, 1962). In applying this model, the position was taken that an anagram constitutes a problem whose stimulus properties evoke a variety of implicit or mediational responses in *S*, and the variables under study—viz., anagram letter order, solution-word frequency, anagram transition probability, and word transition probability—were related to the processes underlying the production of such implicit responses.

More specifically, the use of an S-R mediational model in conceptualizing anagram problem solving involves the following critical assumptions: (*a*) when presented with the stimulus array of letters which constitutes the anagram problem, *S* begins making a series of implicit responses that involve successive rearrangements of the letters into new combinations; (*b*) these rearrangements may involve all of the letters in the anagram or some smaller subset, the smallest possible subset being two letters; (*c*) the production of these letter rearrangements should reflect their frequency of occurrence in the language—e.g., two-letter rearrangements (digrams) with high frequencies of occurrence in the language, and therefore presumably high in *S*'s response hierarchy, should be produced sooner and in larger numbers than digrams with low frequencies of occurrence in the language, and therefore presumably low in *S*'s response hierarchy; and (*d*) *S* solves the anagram when one of these implicit rearrangements matches the word solution, *S* recognizes that a match has occurred, and finally responds overtly with the word solution. The overall solution process then is conceptualized in terms of a series of successive rearrangements or a chain of "linked" implicit responses.

In the foregoing analysis, the third assumption *c* states, in part, that digrams with high frequencies of occurrence in the language should also

TABLE 1

ANAGRAMS, SOLUTION WORDS, ASSOCIATED THORNDIKE-LORGE (T-L) AND DIGRAM FREQUENCIES (DF), AND MEDIAN SOLUTION TIMES (IN SECONDS) FOR HIGH- AND LOW-STORED, DIGRAM-FREQUENCY GROUPS

Anagrams	DF	Solution Words	DF	T-L Freq.	Mdn. Solution Times	
					High	Low
NRCUI	123	INCUR	3	8	184.5	240.0
PMUOI	176	OPIUM	5	7	133.5	240.0
PHMNY	144	NYMPH	8	13	24.0	71.5
RBCSU	192	SCRUB	10	14	10.5	80.5
OAPNR	167	APRON	12	17	83.5	143.5
OCBNA	161	BACON	31	12	56.5	184.0
EODNW	167	ENDOW	43	12	55.0	240.0
LTVIA	205	VITAL	63	21	12.5	29.0
PNCIA	187	PANIC	87	19	44.0	91.0
AWRLB	167	BRAWL	99	5	6.5	14.5
HUOCG	139	COUGH	308	18	10.5	27.0
AEBRL	168	BLARE	321	2	144.0	240.0
SJTUO	135	JOUST	323	12	9.0	143.0
IFNLG	143	FLING	327	15	2.0	6.0
EOKRP	176	POKER	349	4	17.5	28.5

be high in *S*'s response hierarchy. Evidence supporting this assumption (Mayzner & Tresselt, 1962b) showed that *S*s can indeed rank order the frequency with which digrams occur in the language. However, no test has yet been made to determine if *S*s who excel in this ranking task, and therefore presumably have more such digram frequencies "stored" than *S*s who are poor in this task, also excel in solving anagrams. The present study is directed to this issue.

METHOD

Subjects.—The *S*s were 40 students from an introductory psychology course at New York University. No *S* had served in any previous anagram study.

Stimulus material.—The stimulus material consisted of 15 five-letter anagrams typed in caps on white 3 × 5 in. cards. Since previous studies (Mayzner & Tresselt, 1958, 1959, 1962a, 1963) have shown the importance of anagram letter order, solution-word frequency, and anagram and solution-word, transition-probability totals, i.e., summed-sequential digram-frequency values, these variables were controlled for in the present study, as follows: (*a*) all anagrams were presented using hard letter orders; (*b*) all solution words had low and approximately equal Thorndike-Lorge (1944) frequency counts ranging from a low of 2 occurrences per million words to a high of 21 occurrences per million words; (*c*) all anagrams had relatively low and approximately equal summed-sequential digram-frequency values; and (*d*) word solutions were selected so that the associated summed-sequential, digram-frequency values resulted in a group of five solution words with low summed values (ranging from 8 to 17), a group of five solution words with medium summed values (ranging from 31 to 99), and a group of five solution words with high summed values (ranging from 308 to 349).

All anagrams, solution-words, associated summed digram-frequency values, and Thorndike-Lorge frequency values, are presented in Table 1. The summed-sequential, digram-frequency values shown in Table 1 were derived from digram-frequency tables based on a sample of 20,000 words (Mayzner & Tresselt, 1962c). The summed values shown following the anagrams are based on total frequency counts, e.g., for the anagram "NRCUI" the tables give frequency counts of 2, 22, 34, and 65, respectively, for the sequential digrams "NR," "RC," "CU," and

"UI," or a summed value of 123. These frequency counts, 2, 22, 34, and 65, represent the total frequency with which these digrams appear in the 20,000-word sample. In contrast to the anagram values, the sequential digram-frequency values associated with the word solutions are based not on total frequency counts, but rather on frequency counts which take word-length and letter-position variations into consideration. For example, with the word "INCUR," the tables give frequency counts of 1, 0, 0, and 2, respectively, with a summed value of 3, for the sequential digrams "IN," "NC," "CU," and "UR" and are based on the frequencies with which "IN" occurs in the first and second position of five-letter words, "NC" occurs in the second and third position of five-letter words, etc. The rationale for the distinction between total digram-frequency values as applied to anagrams and word-length and letter-position digram-frequency values as applied to word solutions may be found in a recent study by Mayzner and Tresselt (1963).

In brief, differences in solution time between the anagrams should occur only among the three solution-word groups having low, medium, and high summed digram-frequency values, since all other variables, i.e., anagram letter order, solution-word frequency, and anagram digram frequencies, have been held constant. Solution-word, digram-frequency values were specifically varied to determine if the major variable under study, viz., individual differences in stored digram frequencies, might interact with solution-word digram frequencies.

Procedure.—Individual differences in stored digram-frequency values were assessed by a skeleton-word procedure developed in a previous study (Mayzner & Tresselt, 1962b). Using this procedure *S*s were presented in the present study with 60 skeleton-word items. The format for a typical item was as follows: F — — —; R —, I —, L —. The *S* was instructed that each item represented a skeleton word, with the number of dashes representing the number of missing letters, and that his task was to rank order the three letters following the skeleton word in terms of the frequency with which he judged, e.g., that the digrams FR, FI, and FL, would occur in English in the first and second positions of a four-letter word. The 60 skeleton-word items that were used varied in length from three to seven letters and in the letter position of the digrams that were to be judged. Items were scored as correct if *S*s rank ordered the three digrams in correspondence with the word-length and letter-position frequency values of these digrams, as given in the Mayzner and Tresselt tables (1962c).

A total of 106 *S*s, in an introductory psychology class at New York University, were given this 60-item, skeleton-word test. The number of items gotten correctly ranged from a low of 9 to a high of 27; 10 correct represents a chance level of performance. From this distribution of 106 scores, 20 *S*s with scores ranging from 9 to 18 correct (low group) and 20 *S*s with scores ranging from 19 to 27 correct (high group) were selected as representing a low-stored, digram-frequency group and a high-stored, digram-frequency group.

These 40 *S*s were each given in individual testing sessions, several weeks after the skeleton-word test had been given, the 15 anagrams shown in Table 1. Each *S* received the anagrams in a different random order and was allowed 240 sec. to reach a solution. If no solution occurred within 240 sec., *S* was told the solution and then given the next anagram. Solutions were given by *S* verbally (i.e., *S* had no paper or pencil) and solution times were recorded with a stop watch.

RESULTS AND DISCUSSION

Three groups of solution words having high, medium, and low summed digram-frequency values were employed and for each group five solution times were available from each *S* for analysis. Three median solution times were calculated for each *S* for each group of five time scores and served as the criterion measures of performance. Medians were selected rather than means because of skewed distributions and the artificial ceiling imposed by the 240-sec. time limit per anagram.

Means and standard deviations were then calculated, based on the three median criterion time scores, for the three levels of solution-word, digram-frequency values and the two levels of stored digram frequencies, and are presented in Table 2. Application of a 2 × 3 mixed analysis-of-variance design (Edwards, 1960, p.

TABLE 2

Anagram Solution Times (in Seconds) for Solution-Word and Stored Digram-Frequency Groups

Stored Digram Frequencies	Solution-Word Digram Frequencies		
	High	Medium	Low
High			
M	25.20	36.80	80.15
SD	33.25	22.52	68.55
Low			
M	85.25	123.30	143.90
SD	85.28	89.39	86.99

Note.—Based on medians.

227) to the results shown in Table 2 yields an F (1, 38) = 21.90, $p < .001$ for the stored digram-frequency main effect, an F (2, 76) = 7.30, $p < .005$ for the solution-word, digram-frequency main effect, and an F (2, 76) = .46, $p > .20$ for the interaction term.

In addition to the analysis of solution times based on *S*'s performance, an individual word analysis also was performed. In the word analysis, there were available 20 time scores for each word for the high-stored, digram-frequency group and 20 time scores for each word for the low-stored, digram-frequency group. The medians of these two 20 time-score distributions were calculated and are presented in Table 1. In every one of the 15 comparisons possible, the median solution time is less for the high-stored, digram-frequency group than for the low-stored, digram-frequency group and an application of the sign test (Siegel, 1956, p. 68) to these 15 differences yields a $p < .001$.

The present study represents a continuing attempt (Mayzner & Tresselt, 1958, 1959, 1962a, 1962b, 1963, 1964) to test various assumptions of a general S-R mediational model of problem-solving behavior (Kendler, 1961; Kendler & Kendler, 1962) as they apply to the specifics of anagram problem solving. In the earlier work, emphasis was placed on "model" variables which involved systematic variations in the stimulus properties of the anagram and the solution word. In contrast, the present experiment has been concerned with relating possible individual-difference factors in *S* to anagram solution times.

It appears evident from the results of this study that *S*s who have accurately "stored" digram-frequency information, as reflected in high scores on the skeleton-word test, solve anagrams much faster than *S*s who do not have such information in storage, as reflected in low scores on the skeleton-word test. Such individual differences in stored digram frequencies provide still one more source of variance contributing to anagram solution times, which has been identified by the use of an S-R mediational model, and again provides support for the model as a useful way of conceptualizing problem-solving behavior.

REFERENCES

Edwards, A. L. *Experimental design in psychological research.* (Rev. ed.). New York: Rinehart, 1960.

Kendler, H. H. Problems in problem solving research. In, *Current trends in psychological theory: A bicentennial program.* Pittsburgh: Univer. Pittsburgh Press, 1961.

Kendler, H. H., & Kendler, T. S. Vertical and horizontal processes in problem solving. *Psychol. Rev.*, 1962, **69**, 1–16.

Mayzner, M. S., & Tresselt, M. E. Anagram solution times: A function of letter order and word frequency. *J. exp. Psychol.*, 1958, **56**, 376–379.

Mayzner, M. S., & Tresselt, M. E. Anagram solution times: A function of transition probabilities. *J. Psychol.*, 1959, **47**, 117–125.

Mayzner, M. S., & Tresselt, M. E. Anagram solution times: A function of word transition probabilities. *J. exp. Psychol.*, 1962, **63**, 510–513. (a)

Mayzner, M. S., & Tresselt, M. E. The ranking of letter pairs and single letters to match digram and single-letter frequency counts. *J. verbal Learn. verbal Behav.*, 1962, **1**, 203–207. (b)

MAYZNER, M. S., & TRESSELT, M. E. Tables of digram and single-letter frequency counts by word-length and letter-position for a sample of 20,000 words. New York University Library, 1962. (Mimeo) (c)

MAYZNER, M. S., & TRESSELT, M. E. Anagram solution times: A function of word length and letter position variables. *J. Psychol.*, 1963, **55**, 469–475.

MAYZNER, M. S., & TRESSELT, M. E. An exploratory study of mediational responses in anagram problem solving. *J. Psychol.*, 1964, **57**, 263–274.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, Bureau of Publications, 1944.

(Received July 20, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 6, 611-616

PSYCHOPHYSICAL ANALYSIS OF THE ODOR INTENSITY OF HOMOLOGOUS ALCOHOLS[1]

TRYGG ENGEN

Brown University

Change in the human infant's respiration was used as an index of the intensity of homologous alcohols. At 100% concentration intensity decreases but at threshold concentration intensity increases with increasing chain length. Psychophysical scaling with adult Os indicates that the infant results are not inconsistent because the slope of the psychophysical function also decreases with increasing chain length. These findings are pertinent to olfaction as well as psychophysics.

Homologous series of alcohols have been used in the study of olfaction to measure the effect on odor of physical rather than chemical properties, for in ascending such a series there is a gradual change in physical properties such as vapor pressure and solubilities while chemical and structural changes are minimal. Most studies have been concerned with intensity or stimulus efficiency and have involved a variety of organisms and experimental procedures. The relationship between the intensity of an alcohol and its position in the homologous series, however, has not been firmly established (Moulton & Eayrs, 1960). One reason for this is that no complete psychophysical analysis has been made for any of the organisms used. Typically investigators have determined only thresholds or some other suprathreshold concentration for each alcohol. One index of this sort has then been used as the sole measure of the intensity of the odor, although it has long been suspected that threshold data may not predict suprathreshold data and vice versa (Kruger, Feldzamen, & Miles, 1955).

The present paper presents both threshold and suprathreshold data on the human infant, a very good psychophysical *S* because of the simplicity of the response which can be obtained from it. Supplementary psychophysical scales were also obtained from the human adult.

METHOD

Subjects.—The *S*s in the first two basic experiments were 125 apparently normal human infants. Their mean age was 58 hr. with a range of 20–105. There were more girls than boys (roughly 2:1) because boys were often circumcised during the period when they were normally available for the experiment. The *S*s had not participated in any previous olfaction experiments. Forty-two graduate students and staff members of the psychology department served as *O*s in the scaling experiments.

Apparatus and materials.—The infant experiments utilized a stabilimeter crib designed for measurement of various responses (Lipsitt & DeLucia, 1960). In this case breathing was observed with a Phipps and Bird infant pneumograph attached around the abdomen and recorded on a Grass polygraph (Model 5A) run at 5 mm/sec.

The odorants were a series of homologous alcohols ranging from ethanol to decanol, or 2–10 carbon atoms, except nonanol which was not available; the alcohols were supplied by Eastman Organic Chemicals and were redistilled. Diethyl phthalate produced by

---

[1] The research is supported by a grant (GB-724) from the National Science Foundation. Lewis P. Lipsitt kindly gave us the opportunity to conduct the infant research through his United States Public Health Service Grant (NB 04268). William S. Cain carried out the adult-scaling experiment, and Dorothy Westlake assisted in the infant experiments.

Fritzsche Brothers was used as diluent in the infant threshold and adult-scaling experiments after comparisons with distilled water.

One cubic centimeter of each odorant was prepared in a 10 × 74 mm. Pyrex test tube stopped with a cork wrapped in aluminum foil. The odorants were presented to *S*s on cotton wrapped around a glass rod attached to the cork. The cotton was kept in the test tube and partially immersed in the liquid odorant when not in use.

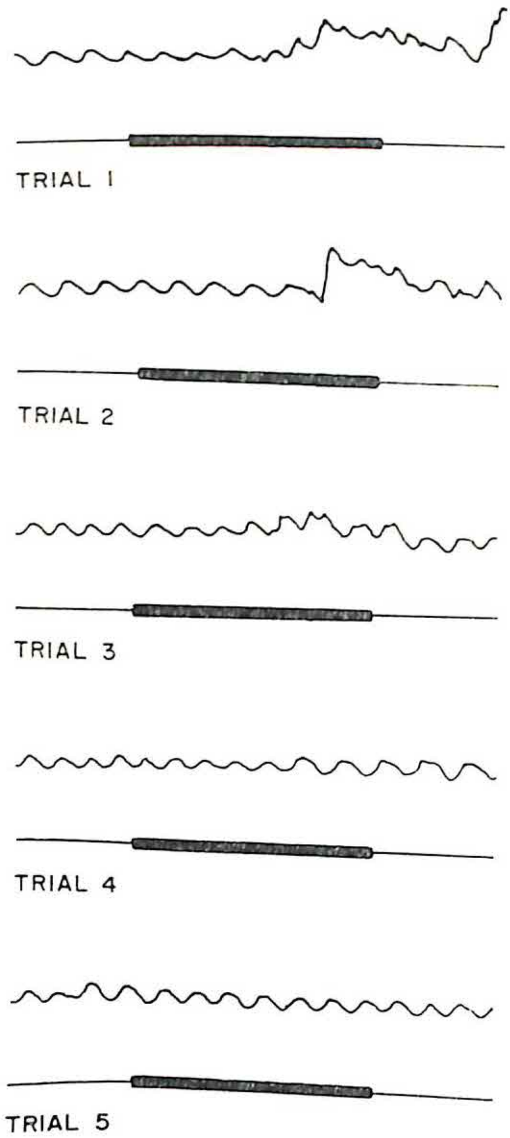


FIG. 1. Sample breathing record showing 10-sec. pen marks for five consecutive presentations of a 100% concentration of pentanol. (The event marker moved up with inspiration and down with expiration. Responses, as defined in text, were obtained on Trials 1, 2, and 3 but not on Trials 4 and 5. The response usually occurs after about three inspirations.)

Procedure.—The experiment was performed at the Providence Lying-In Hospital. The temperature in the infant laboratory was maintained close to 78° F. All infants were tested individually between 10:00 and 11:00 A.M., approximately 1 hr. after feeding at the time when they are relatively quiet and sleepy. The infant was placed on its back in the stabilimeter crib and the pneumograph was attached. Testing would start when the 2 *E*s judged the infant to be asleep, i.e., eyes closed, regular breathing, and no activity indicated on the polygraph. Occasionally an infant could not be used because of failure to quiet down according to this criterion. The infants were assigned at random to the conditions of each experiment.

One *E* presented the stimuli to *S*, while the other *E* entered pen-marker signals on the polygraph paper to record the start and removal of each stimulus. A trial consisted of holding the cotton about 5 mm. below *S*'s nostrils for 10 sec. Sometimes *S* responded by moving its head and this ended the trial. A trial was presented every 30 sec., although at times a longer interval was required before the infant settled down again to the pretrial criterion. A response was defined as any disruption in *S*'s normal breathing pattern by increase or decrease in amplitude and rate (Engen, Lipsitt, & Kaye, 1963). More than one form of change often occurred in the same record, but only one was required for the present definition of response. Since tension on the pneumograph strap varied between *S*s, amplitude cannot be quantitatively specified, but rate was normally 30–50 cycles/min in the sleeping infant.

One group of 10 infants was run for five trials but with only the odorless diluent on the cotton rather than alcohol. For only three of those 10 *S*s a change in the breathing record was obtained on one of the five trials, a response frequency safely below that of experimental *S*s described below.

RESULTS

Experiment I: Undiluted alcohol.—Five trials were administered to each infant with an alcohol at 100% concentration following the procedure used earlier (Engen et al., 1963) with other odorants. A sample record from one infant is presented in Fig. 1. The score for each infant was the number

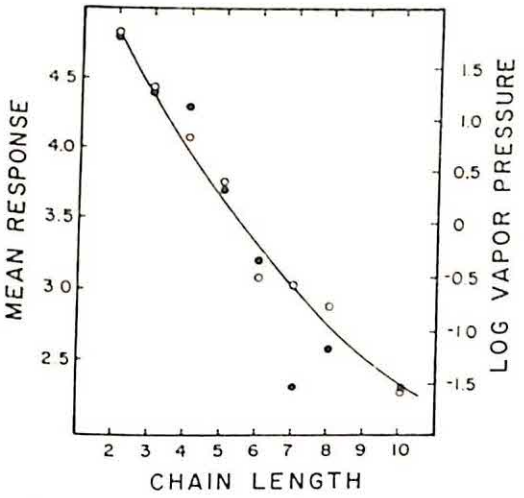


FIG. 2. Mean response (●) and log vapor pressure (o) at 25° C. as function of alcohol chain length. (The curve was fitted by eye.)

of responses given on those five trials. There were 10 different *S*s in each group, and each group was presented with only one of the eight alcohols. The results are presented in Fig. 2 which shows clearly that the mean number of responses is a decreasing function of the number of carbon atoms in the chain. It is not known why the point for heptanol (7) deviates from the otherwise monotonic function. The *SD* for each point plotted in Fig. 2 is about 1.2. There is a suggestion that variability of response is proportional to chain length, or inversely proportional to intensity. Figure 2 also presents the vapor pressures of the alcohols, which show a decline with chain length very similar to the infant's response.

The results of this experiment are in agreement with results reported by Kruger et al. (1955) with adult human *O*s who matched the intensities of each of these alcohols to one of a range of concentrations of heptanal. They conclude:

> Within most of the range of alcohols studied it was found that olfactory intensity decreases with longer chain molecules. This trend appears, however, to be reversed after correcting for vapor-pressure [p. 394].

However, most of the research in this area has shown that threshold concentrations of these alcohols for flies (Dethier & Yost, 1952) and rats (Moulton & Eayrs, 1960) as well as human *O*s (Skramlik, 1948) decrease as a function of chain length without correcting for vapor pressure. In other words, on this basis the longer chain alcohols are more intense or efficient stimuli than the shorter chain

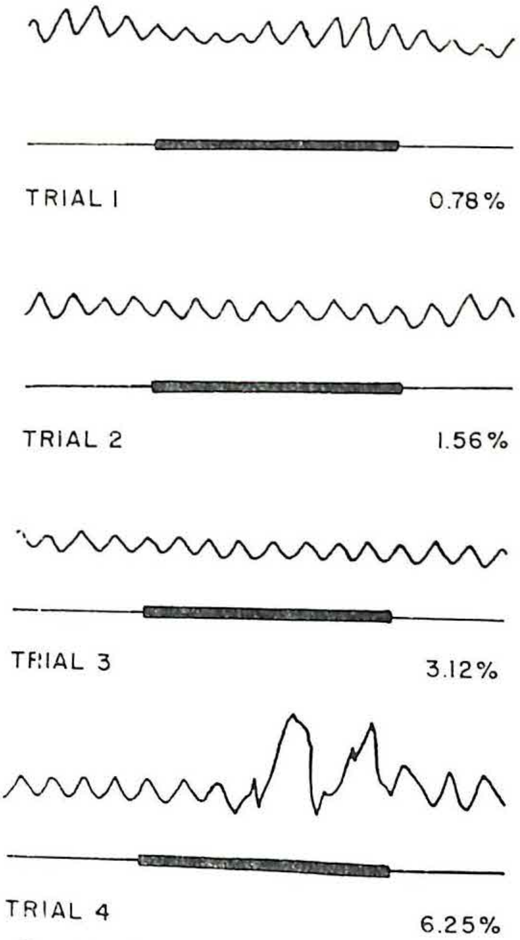


FIG. 3. Sample breathing record showing 10-sec. pen marks for consecutive trials with a geometrically increasing concentration of pentanol. (Threshold as defined here is reached on Trial 4 with 6.25% concentration. The record for Trial 1 is doubtful but is not considered a threshold response since there is no change in breathing pattern during the next two trials with increased concentration.)

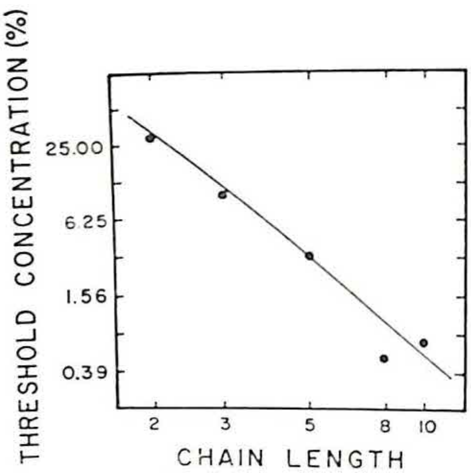


FIG. 4. Threshold concentration as function of alcohol chain length. (The line was fitted by eye.)

alcohols. The next experiment therefore concerned threshold determinations in the human infant.

Experiment II: Alcohol detection.—The experiment used the method of limits applied in earlier research (Lipsitt, Engen, & Kaye, 1963). A $\log_2$ concentration dilution series of the odorant was prepared by pipette. On the first trial the infant was presented a concentration well below threshold and the concentration of the odorant was doubled in successive trials until the infant responded. The concentration used on that trial defines threshold in this case. A sample record is shown in Fig. 3. In other respects the procedure was the same as above. Thresholds were obtained for ethanol, propanol, pentanol, octanol, and decanol for five groups of five *S*s. The mean thresholds are plotted in Fig. 4 and are also a decreasing function of chain length. The *SD* is approximately one half step on the geometric-concentration scale for each group. When water was used as diluent, no appreciable change was noted for propanol in either mean or *SD* but for octanol the threshold was 1.2 $\log_2$ steps higher and the *SD* was 1.36 as compared with .49 for dilution with diethyl phthalate. This latter finding is not surprising, considering the low solubility of octanol in water. The distribution of alcohol in the water was obviously uneven; in fact the solution was difficult to prepare for it clogged the pipette. In an earlier experiment with adults as *O*s, and benzyl benzoate as diluent, thresholds were also less reliable (*SD*s of about 2.5) and revealed no clear relationship with chain length (Engen, 1963).

It is now apparent that the infant has produced two sets of conflicting data on alcohol intensity. As measured by responsiveness to 100% concentrations, the shorter the chain the stronger the alcohol, but as measured by threshold concentrations, the longer the chain the more efficient the alcohol. It was not possible to determine vapor pressures and thermodynamic activities (Dethier & Yost, 1952) in the present case, but the present threshold data agree in a general way with the analysis of Moulton and Eayrs (1960) who conclude, on the basis of their own and other workers' published results, that threshold expressed as pressure appears to be a linear function of log chain length.

A problem remaining now is to determine the magnitude of response for solutions between threshold concentration and 100% concentration of the alcohols with a simple and direct approach. With the present infant apparatus we were able to determine only the presence or absence of response. Therefore, we turned to the adult human *O* with this scaling problem. Since the same apparent conflict in results as was found in the present data is also provided by the data of Kruger et al. (1955) on 100% concentrations and the data of Skramlik (1948) on threshold, it is not unreasonable to turn to adult *Ss* for this further scaling.

Experiment III: Scaling.—This experiment applied the method of magnitude estimation without a prescribed modulus (Stevens, 1956) to a range of concentrations of propanol and octanol, selected to represent short vs. long chain lengths. Seven concentrations of each alcohol were used. These concentrations varied from 100% to 1.56% and therefore roughly covered the range used with infants. Thus the same stimuli are used, but in this experiment response magnitude is evaluated by numerical judgments of intensity from sniffing the cotton. This procedure has been described earlier (Engen & Lindström, 1963).

The instructions read to *O* stressed that he should judge intensity with a constant sniff and ignore all other attributes of the odor. He had no knowledge of the infant data; nor was he informed about the stimuli. His task was to assign any number he felt appropriate to the intensity of the first stimulus, and assign numbers to subsequent stimuli presented 1 min. apart such that the numbers were proportional to the intensity of the odors on a ratio scale. The *Os* were randomly divided into two groups of 11. Each concentration was presented twice, and a different irregular order was used for each *O*. Geometric means were computed for the numbers assigned to each stimulus and these are plotted against concentration in Fig. 5 in log coordinates. The lines are fitted by the method of least squares and indicate that both sets of data conform to the power law (Engen & Lindström, 1963). However, there is a surprising difference in the slope (.13 vs. .42), the parameter which generally specifies the input-output characteristic of a sense modality (Stevens, 1957). A preliminary version of the present experiment with two similar groups of 10 *Os* indicated the same difference in slope but overall octanol was judged somewhat weaker.

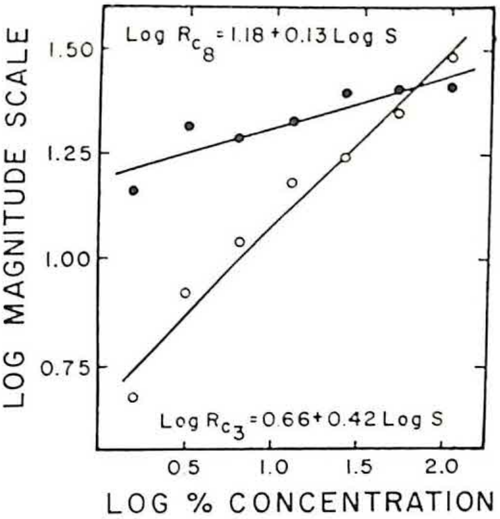


FIG. 5. Log magnitude-estimation scale as function of concentration of octanol (C_8) and propanol (C_3).

Propanol is stronger than octanol at 100% concentration, but while the intensity of propanol decreases rapidly with dilution and approaches threshold at 1.56%, the intensity of octanol changes very little when diluted in the same manner. In fact, several *Os* assigned a higher number to the 1.56% than they did to the 100% concentration of octyl alcohol, an error rarely made even with two adjacent concentrations of propanol.

DISCUSSION

It can now be seen that the results of Exp. I and II are consistent, assuming that the slope of the psychophysical function decreases with chain length as in Exp. III. Thrsehold concentrations, intensity of undiluted alcohols, and the degree to which dilution affects this intensity, *all* decrease with increasing chain length. It is clear that one cannot generalize about the intensity of a whole range of effective stimulus concentrations with information about only thresholds or other selected values. This kind of prediction must be made on the basis of the appropriate psychophysical scale, which in the present case ought to be determined for the infant. However, the present psychophysical analysis points to the similarity of certain aspects of olfaction in different organisms despite structural and developmental differences.

The results on undiluted alcohols verify the importance of vapor pressure and it has been shown that thermodynamic activity is involved (for review see Moulton & Eayrs, 1960). However, the different slopes of the otherwise similar psychophysical scales obtained for propanol and octanol suggest that something more than vapor pressure determines the magnitude of response. The receptors respond to variation in concentration of the two alcohols in the same manner (power function) but not to the same degree (slope), because dilution has a much greater effect on propanol than octanol. This latter difference might be akin to the binding force of molecules with the receptors described in Beidler's taste theory (1961).

Of course it is not clear what kind of receptors are stimulated with the present methodology. Tucker (1963) has shown with simultaneous electrophysiological recording from the primary olfactory nerve fiber and the trigeminal nerve in the rabbit that at high concentration the trigeminal nerve probably responds before the olfactory nerve to short chain alcohols. Generally, however, a higher threshold concentration of an odorant is required to stimulate the trigeminal nerve than the olfactory nerve. It is possible, e.g., that the infant's responses to at least some undiluted alcohols depend largely on the trigeminal nerve endings, while responses to threshold concentrations depend more on olfactory receptors. Only further simultaneous recording from both nerves could settle this problem, which may also be the basis for qualitative variation of the smell of alcohols correlated with chain length (Engen, 1964). Nevertheless, progress has been made on the "outer" psychophysics of homologous alcohols.

REFERENCES

Beidler, L. M. Mechanisms of gustatory and olfactory receptor stimulation. In W. A. Rosenblith (Ed.), *Sensory communication.* New York: Wiley, 1961. Pp. 143–157.

Dethier, V. G., & Yost, M. G. Olfactory stimulation of blowflies by homologous alcohols. *J. gen. Physiol.*, 1952, **35**, 823–839.

Engen, T. Cross-adaptation to the aliphatic alcohols. *Amer. J. Psychol.*, 1963, **76**, 96–102.

Engen, T. Psychophysical scaling of odor intensity and quality. *Ann. N. Y. Acad. Sci.*, 1964, **116**, 504–516.

Engen, T., & Lindström, C. O. Psychophysical scales of the odor intensity of amyl acetate. *Scand. J. Psychol.*, 1963, **4**, 23–28.

Engen, T., Lipsitt, L. P., & Kaye, H. Olfactory responses and adaptation in the human neonate. *J. comp. physiol. Psychol.*, 1963, **56**, 73–77.

Kruger, L., Feldzamen, A. N., & Miles, W. R. Comparative olfactory intensities of the aliphatic alcohols in man. *Amer. J. Psychol.*, 1955, **68**, 386–395.

Lipsitt, L. P., & DeLucia, C. A. An apparatus for the measurement of specific response and general activity of the human neonate. *Amer. J. Psychol.*, 1960, **73**, 630–632.

Lipsitt, L. P., Engen, T., & Kaye, H. Developmental changes in the olfactory threshold of the neonate. *Child Develpm.*, 1963, **34**, 371–376.

Moulton, D. G., & Eayrs, J. T. Studies in olfactory acuity. II. Relative detectability of n-aliphatic alcohols by the rat. *Quart. J. Psychol.*, 1960, **12**, 99–109.

Skramlik, E. von. Über die zur minimalen Erregung des menschlichen Geruchs bzw. Geschmackssinnes notwendigen Molekülmengen. *Pflüg. Arch. ges. Physiol.*, 1948, **249**, 702–716.

Stevens, S. S. The direct estimation of sensory magnitudes—loudness. *Amer. J. Psychol.*, 1956, **69**, 1–25.

Stevens, S. S. On the psychophysical law. *Psychol. Rev.*, 1957, **64**, 153–181.

Tucker, D. Olfactory, vomeronasal and trigeminal receptor responses to odorants. In Y. Zotterman (Ed.), *Olfaction and taste.* Oxford: Pergamon, 1963. Pp. 45–69.

(Received July 20, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 6, 617-625

CUTANEOUS SOUND LOCALIZATION [1]

GEORGE A. GESCHEIDER [2]

University of Virginia

Cutaneous sound localization when stimuli were delivered to the skin through a pair of vibrators was compared with auditory localization when stimuli were presented through earphones. Auditory localization was more precise for random noise bursts than for low-frequency tones. Cutaneous localization, however, was as accurate for the tone as for the noise stimuli. Comparisons between modalities revealed that cutaneous localization of low tones was a great deal more precise than auditory localization of low tones. Localization of noise bursts, however, was slightly more accurately performed by the ears than by the skin, as was the case with 1-msec. clicks. Independent manipulation of intensity- and temporal-difference cues revealed that auditory localization was influenced by both types of cue. Cutaneous localization was found to depend mainly on intensity differences. Small time-difference effects were observed, however, and were found to depend on the locus of stimulation.

One approach to sensory experimentation is the investigation of phenomena common to the various sensory modalities. Much of the work of Békésy, e.g., is typical of this approach. He has reported that with training human *S*s can accurately localize sound sources when stimulation is exclusively cutaneous (Békésy, 1955). In this experiment, the two forearms were stimulated by means of cochlear models. Acoustic stimuli activated two spatially separated microphones, and the signals were amplified to drive the two cochlear models. This arrangement presented to the two forearms the temporal and intensive differences in stimulation known to play important roles in auditory localization. The experiments described in this report were designed to compare the precision of cutaneous and auditory localization and to provide evidence concerning cue utilization in both modalities.

EXPERIMENT I

Since the ear localizes noises more precisely than pure tones (Rayleigh, 1877; Stevens & Newman, 1936), an attempt was made to compare the localization of these two kinds of stimuli by the skin.

Auditory localization is known to depend upon cues produced by temporal and intensive differences in stimulation between the two ears, the relative importance of each being dependent upon specific stimulus conditions. Cue utilization for both cutaneous and auditory localization was compared by measuring the effects of providing *S*s with conflicting temporal and intensive cues.

Method

Subjects.—Eight *S*s, male students at the University of Virginia, were employed.

Apparatus.—Stimuli were presented through a 4-in. dynamic speaker located a few

---

[1] This investigation was supported in part by Public Health Service Research Grant NB04177 from the National Institute of Neurological Diseases and Blindness. It is based on a dissertation submitted in partial fulfillment of the requirements for the PhD degree at the University of Virginia. The author would like to thank Raymond C. Bice and John F. Hahn for their help throughout the project.

[2] Now at Hamilton College.

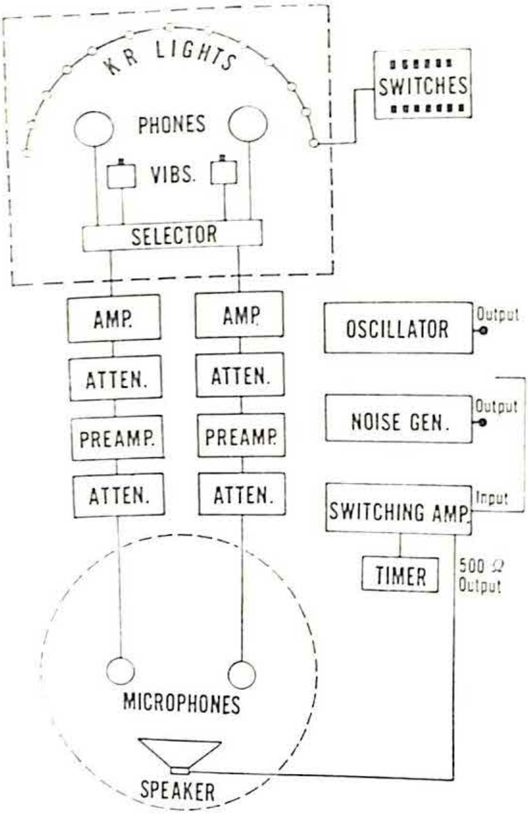


FIG. 1. Diagram of apparatus for Exp. I.

inches from the inner wall of a circular enclosure $7\frac{1}{2}$ ft. in diameter and 7 ft. high made of 1-in. fiberglass. Two low-impedance dynamic microphones were mounted 27 cm. apart on a horizontal bar located in the center of the enclosure 80 cm. from the speaker. By rotation of this horizontal bar the microphone axis could be rotated in relation to the fixed speaker location. The rotation in degrees was read from a scale below the horizontal bar. A cylindrical roll of fiberglass between the microphones served as an analog of the human head, producing a sound shadow for the higher frequencies.

Acoustic stimuli produced by the speaker activated the two microphones, the signals from which were separately amplified and controlled by solid-state preamplifiers, amplifiers, and attenuators (Fig. 1).

In a separate cubicle the electrical signals were transduced to mechanical energy by either a pair of Goodmans V-47 vibrators or by earphones. The vibrators were used to deliver stimulation to the index fingertips of the two hands. It can be seen in Fig. 1 that each earphone or vibrator was driven by an independent microphone-preamplifier-amplifier system. Thus, when the sound-source location was varied, this arrangement permitted preservation between the two channels of temporal and intensive differences.

The *S* was seated in the dark behind a semicircular board 72 in. in diameter. The vibrators were mounted below the edge of the board at approximately knee level. In response to auditory or cutaneous stimuli, a 33-in. luminous pointer could be rotated by *S* on a pivotal point at the center of the semicircle. A scale with 1° units from 90° left to 90° right was painted on the board so that a displacement of the pointer could be read and reported by *S*. Located at every 15° on the circumference of the board were dim red lights to provide knowledge of results (KR) after each judgment. Thirteen microphone positions corresponding to the 13 light positions were used during the experiment.

The stimuli used in Exp. I were noise bursts and low-frequency tones of 1-sec. duration. The noise bursts were produced by a Grason-Stadler Model 901A noise generator. The noise had a speech-frequency bandwidth of equal energy per cycle, filtered to fall off at 12 db. per octave above 1,000 cps. The frequency response of the vibrators and of the sense of touch both required the selection of a tone of relatively low frequency, and it was determined empirically that a frequency of 187 cps produced minimal distortion by reflection from the walls of the room. The tones were produced by a General Radio 913-C beat-frequency oscillator. The output of the oscillator or noise generator was controlled by a switching amplifier to eliminate transients. Stimulus duration was controlled by a timer.

Procedure.—Each *S* was observed under all modality stimulus conditions. A 4 × 4 balanced Latin square containing conditions of cutaneous and auditory localization of tones and noise bursts determined the order of conditions for each *S*. Three experimental sessions were conducted successively under each of the four conditions. Five trials were administered during each of the first two sessions and two trials during the third session. Each trial consisted of a random presentation of the 13 sound-source locations which varied in azimuth from 90° to the left to 90° to the right.

At the beginning of each session the sound source was set at 0° azimuth and the intensity level of the left channel was adjusted until *S* reported the sensation to be localized at 0° azimuth. The obtained attenuator settings were then maintained throughout the session.

The following occurred with each stimulus presentation: (*a*) a small light mounted on the pointer was turned on as a ready signal and 2 sec. later the stimulus was presented; (*b*) after the pointer was adjusted by *S* the appropriate KR light was presented; and (*c*) *S* verbally reported the pointer-scale reading in degrees.

Following Trials 11 and 12 of the third session for a given condition, the relative importance of intensive and temporal cues was investigated. The voltage at each earphone or vibrator produced by stimulus presentation was measured for each of the 13 sound-source locations. Voltage differences between the two transducers represented the intensive cues associated with various locations. It was expected that if temporal cues were not used by *S*, setting the sound source at 0° azimuth (zero time difference) and directly producing the voltages associated with the various sound-source locations (Cond. I) would lead to no significant change in judgments from those observed on Trials 11 and 12 (Cond. IT). If, however, temporal cues are important, it was expected that these operations would result in a shifting of all judgments toward the 0° azimuth.

Results and Discussion

Precision of localization.—The average error (discrepancy, ignoring sign, between pointer setting and sound-source location) in degrees was employed as a measure of the precision of localization. Since auditory and cutaneous localization improved equally with practice, statistical analysis was performed on the data combined over the 12 trials.

Consistent with previous findings, auditory localization of noise bursts (average error 12.8°) was significantly more precise than localization of low-frequency tones (average error 24.8°), F (1, 18) = 128.18, $p < .01$. Stevens and Newman (1936) have proposed that pure tones provide only a single cue to an *S* (phase differences for low-frequency tones and intensity differences for high-frequency tones) and thus are not as precisely localized as more complex stimuli which provide both types of cue. The major cue for localizing low-frequency tones such as those used in Exp. I is known to be the binaural phase difference. In the present experiment, however, there were also intensity differences for the low-frequency tone employed (7 db. at 90° azimuth), indicating that conditions were different from those that exist when measurements are made between the ears of an actual human head in a reflection-free sound field (Sivian & White, 1933). It is likely that auditory tone localization was influenced by both intensity-difference and phase-difference cues, but the presence of reflections led to distortion of the phase difference, thus limiting localization accuracy. Since a regular relationship existed between the voltage differences between the transducers and the sound-source locations, distortion of intensity cues could not account for the poor localization of tones.

The average error for cutaneous localization was not significantly different for noise (15.6°) and tone (14.3°) stimuli, F (1, 18) = 1.37, $p > .05$. The results of independently manipulating temporal- and intensive-difference cues revealed that the skin is responsive to intensity differences provided by noise and tone stimuli but is little influenced by temporal cues, whereas both cues affected auditory localization (described below). Thus, the distortion of the phase-difference cue could not interfere with cutaneous tone localization as it did in auditory localization.

Cutaneous localization of tone stimuli was significantly more precise than auditory tone localization, F (1, 18) = 86.80, $p < .01$. Noise bursts, however, were more precisely localized by the ears than the skin, F (1, 18) = 10.09, $p < .01$.

Cue utilization.—Presented in Fig. 2 is the mean pointer displacement as a function of sound-source location

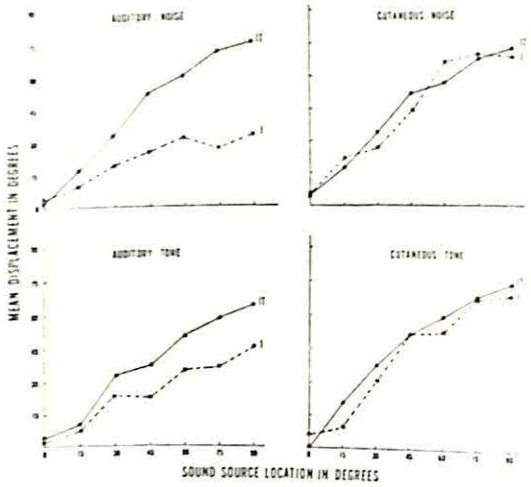


FIG. 2. Mean pointer displacement as a function of sound-source locations 0°–90° (averaged over left and right quadrants) for IT and I cue conditions for auditory and cutaneous localization of noise bursts and tones.

when both intensive and temporal differences were varied with the sound source (IT) and when intensive differences were varied with temporal differences always at zero (I). For auditory noise localization it is apparent that holding time differences constant at zero produced significant shifts toward 0° azimuth, $F\ (1, 18) = 153.48$, $p < .01$. In contrast, significant shifts in localization were not observed for cutaneous noise localization, $F\ (1, 18) = 1.00$, $p > .05$. Thus, auditory noise localization was influenced by both intensive and temporal cues while cutaneous localization was not affected by the temporal-difference cue but rather was entirely dependent upon intensive-difference cues.

The data for tone localization reveal a similar though less dramatic story. Significant IT-I shifts were observed for auditory localization, $F\ (1, 18) = 41.94$, $p < .01$, indicating a dependence on both phase-difference and intensity-difference cues. Since it is assumed that low-frequency tones provide only phase-difference cues, much larger shifts in localization were expected than those observed in the present experiment. However, as stated earlier, acoustic conditions differed somewhat from those measured for a human head in a reflection-free sound field. Cutaneous localization of tones resulted in a very small but statistically significant IT-I shift, $F\ (1, 18) = 4.95$, $p < .05$. This shift was significantly smaller than the one observed for hearing, $F\ (1, 18) = 9.00$, $p < .01$. Thus, as was found for noise localization, the skin was much less sensitive to small temporal differences in stimulation than were the ears.

EXPERIMENT II

Experiment II was conducted to compare the precision of auditory and cutaneous localization of clicks. It was also hoped that the use of clicks would provide more data relevant to the problem of cue utilization in cutaneous localization. The results of Exp. I seemed to indicate that the skin is relatively insensitive to the temporal cues provided by a noise or low-frequency tone. However, the short duration and abrupt onset of click stimulation offer optimal conditions for the occurrence of temporal effects on the skin, and therefore if temporal cues influence cutaneous localization at all it was felt that the use of clicks would reveal these effects.

To increase stimulus control, a sound field was not employed in Exp. II. The cues were manipulated directly by electronically generating two separate clicks of desired temporal and intensive characteristics and delivering them independently to the two transducers.

Method

Subjects.—Four University of Virginia psychology graduate students served as *S*s.

Apparatus.—A set of Tektronix 160-series components consisting of a power supply, waveform generator, and two pulse generators was used to generate clicks of variable amplitude, duration, and time of onset. Signals from two pulse generators served to trigger two transistorized switches that applied 6-v. dc power supplies to the transducers. Intensity-difference cues were varied by increasing or decreasing the resistance in the two circuits. The intensity level of stimuli originating from 0° azimuth was approximately 60 db. SL and 40 db. SL for the auditory and cutaneous stimuli, respectively.

As in Exp. I, stimuli were presented from 13 sound-source locations. The time-difference values for these locations ranged from 0 to .653 msec. and corresponded to values determined between the ears of a human head in an actual sound field (Woodworth & Schlosberg, 1955). Since intensity-difference values were not available they had to be determined through the use of a psychophysical procedure. A series of clicks having a binaural time difference associated with the 45° sound-source location was presented by earphones to a single *S*, and the binaural intensity difference was varied until *S* reported the sound image to be localized at 45°. The psychophysically obtained intensity difference was approximately two thirds as large as the comparable difference measured between the microphones in the sound field of Exp. I. Furthermore, this reduction value of two thirds was found to hold for all 13 sound-source locations. Intensity-difference and time-difference values for sound-source locations from 0° to 90° to the left and right are presented in Table 1.

Because it was felt that temporal effects on the skin might not be found for time differences within the range produced by an actual sound source (0 to .653 msec.), time differences of greater magnitude were also used (1, 2, 4, and 6 msec.).

Since different transducers were used for the ears and the skin it was essential that an electronically generated time delay produce the same delay in mechanical stimulation between the earphones as between the vibrators. Observation on an oscilloscope of the response patterns of the vibrators and the earphones revealed this to be the case. Furthermore, when the vibrators were placed one next to each ear to deliver an audible click, a time delay of between 1 and 2 msec. displaced the sensation maximally to the ear first stimulated in exactly the same fashion as did the earphones.

Procedure.—Auditory and cutaneous localization were observed on alternate days for a total of four sessions each. All combinations of three cue conditions and 13 sound-source locations were presented once in a randomized order within a session. The three cue conditions consisted of: presenting together the intensity and time differences associated with the various sound-source locations (IT); holding constant at zero the time difference while intensity-difference cues were varied (I); and holding constant at zero the intensity difference while time-difference cues were varied (T).

Knowledge of results was provided only after localization of IT stimuli. It was felt that if IT performance was maintained at a high level throughout the session, I and T shifts in localization could clearly be attributed to cue changes.

Results and Discussion

Precision of localization.—The average error for the localization of clicks was significantly smaller for the ears (8.0°) than for the fingertips (10.3°), $F(1, 75) = 6.46$, $p < .01$.

Cue utilization.—In Fig. 3 mean pointer displacement from 0° azimuth as a function of sound-source location for the IT, I, and T cue conditions for

TABLE 1

INTENSITY-DIFFERENCE AND TIME-DIFFERENCE VALUES FOR SOUND-SOURCE LOCATIONS FROM 0° TO 90° AZIMUTH

	Sound-Source Location in Degrees						
	0	15	30	45	60	75	90
Intensity Difference (db.)	0.00	2.49	4.28	5.82	7.40	8.60	8.66
Time Difference (msec.)	.000	.132	.260	.379	.486	.578	.653

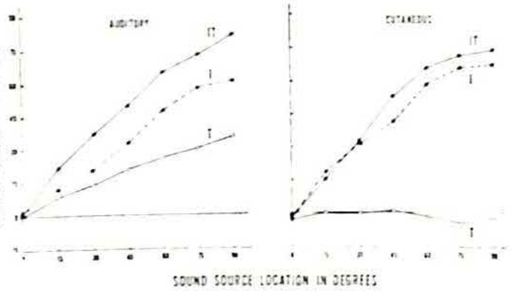


FIG. 3. Mean pointer displacement as a function of sound-source locations 0°–90° (averaged over left and right quadrants) for IT, I, and T cue conditions for auditory and cutaneous click localization.

auditory and cutaneous localization is presented. As was the case in Exp. I for noise and tone stimuli, variation of intensity-difference cues with zero time difference (I) produced significantly less auditory displacement than when intensity differences and time differences varied together (IT), $F\ (1, 105) = 109.09$, $p < .01$. Thus, auditory localization of clicks was influenced by both binaural time and intensity differences.

When binaural time differences were varied and binaural intensities set at equal loudness, significant displacements from 0° azimuth were observed, $F\ (1, 105) = 440.39$, $p < .01$. The displacement of the T curve, however, was significantly less than that of the I curve, $F\ (1, 105) = 111.89$, $p < .01$. Thus the intensity differences used in this experiment had a greater effect than did time differences.

In Fig. 3 the comparable results for cutaneous localization are also presented. Variation of intensity differences with time differences at zero (I) produced significantly less displacement than the IT condition, $F\ (1, 105) = 10.19$, $p < .01$. The difference, although slight, might be interpreted as evidence that cutaneous localization of clicks depends, at least to a small extent, upon time-difference cues. When time differences were varied between stimuli of equal subjective intensity (T), however, displacement from 0° azimuth failed to be statistically significant, $F\ (1, 105) = 1.26$, $p > .05$.

It is evident that cue utilization for auditory and cutaneous localization of clicks is quite different. Auditory localization depends on both binaural intensity differences and binaural time differences but cutaneous localization appears to be dependent, almost completely if not entirely, upon intensity differences between the fingertips.

The relationship between displacement of the pointer and time differences of 0 to 6 msec. is presented in Fig. 4. In auditory localization the amount of displacement is a regular linear function of increasing binaural time differences up to a point somewhere between .653 and 1.000 msec. with less rapid increases in displacement with further increases in the time difference. At values greater than 2 msec., double sound images,

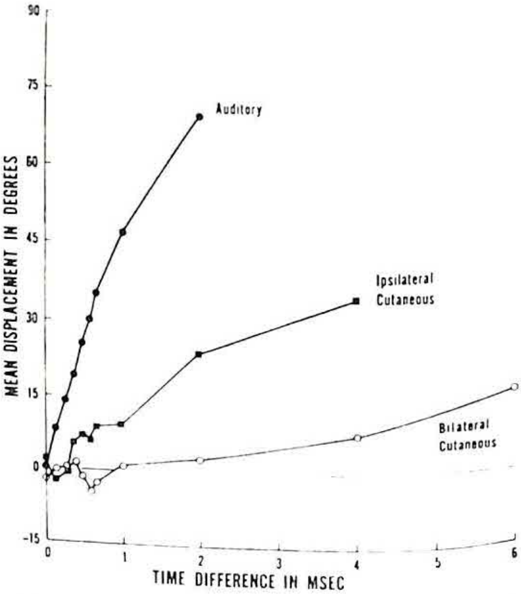


FIG. 4. Mean pointer displacement as a function of time difference for auditory, ipsilateral cutaneous, and bilateral cutaneous conditions.

one in each ear, were reported, and this lack of fusion of the images prevented precise localization. The lowest curve in Fig. 4 (Bilateral Cutaneous) represents the relationship between time differences in stimulation and cutaneous localization in Exp. II. It was reported above that time differences corresponding to those normally encountered in hearing in a free field (0 to .653 msec.) have little effect on cutaneous localization. However, as time differences increased from 1 to 6 msec., there was a systematic increase in the amount of displacement. The *F* ratios computed for the individual values revealed that only time differences as great as 4 msec., $F\ (1, 9) = 9.33$, $p < .05$, significantly influenced cutaneous localization when corresponding bilateral fingertips were stimulated.

EXPERIMENT III

Of primary concern in Exp. III was the arrangement of conditions to demonstrate greater effectiveness of time-difference cues in cutaneous localization. If time differences in stimulation influence localization by producing neural inhibition, and more inhibition occurs when the stimulated areas are closely interconnected neurologically, then it might be expected that stimulation of the fingertips of the same hand (ipsilateral stimulation) would result in greater time-difference effects than stimulation of the index fingertips of the two hands (bilateral stimulation).

Method

Subjects.—The four *S*s who participated in Exp. II were also used in Exp. III.

Apparatus.—The apparatus was identical to that used in Exp. II. The only modification was to move the vibrators close enough together to stimulate the index and ring fingers of the left hand.

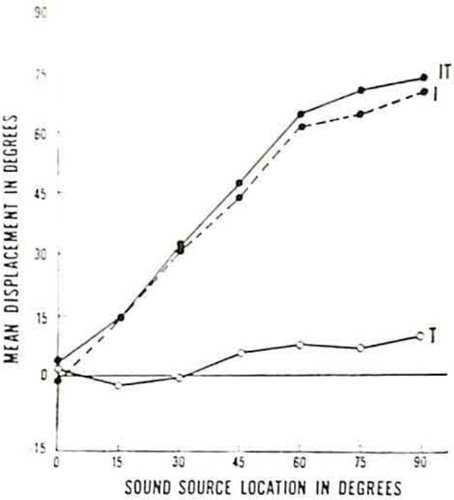


FIG. 5. Mean pointer displacement as a function of sound-source locations 0°–90° (averaged over left and right quadrants) for IT, I, and T cue conditions for ipsilateral-cutaneous click localization.

Procedure.—As in Exp. II, the effects of IT, I, and T cue conditions and the larger time-difference values of 1 to 6 msec. were observed within each experimental session. Two sessions were conducted on successive days.

Results and Discussion

Mean pointer displacement is plotted as a function of the location of the sound source for the three cue conditions (IT, I, and T) in Fig. 5. A significant shift in the direction of 0° azimuth was observed when time differences were held constant at zero (I), $F\ (1, 51) = 19.03$, $p < .01$, indicating that time differences as small as or smaller than .653 msec. may affect cutaneous localization. Consistent with this finding was the significant amount of displacement produced by introducing these small time differences when intensity differences were zero (T), $F\ (1, 51) = 11.27$, $p < .01$.

When clicks with larger time delays between the vibrators (1–6 msec.) were presented, increasing displacements were observed. As seen in

Fig. 4, a maximum displacement of 35° was produced by a 4-msec. time difference. At 6 msec., however, two successive sensations were reported and localization was impossible. It is clearly evident that time differences have greater effects in ipsilateral than in bilateral cutaneous stimulation. The cutaneous effect, however, is still considerably smaller than that observed for hearing.

EXPERIMENT IV

The results of Exp. I, II, and III seemed to indicate that time-difference cues have considerably more influence on auditory than on cutaneous localization. Experiment IV was an attempt to extend this finding beyond the specific measurement technique used in the previous three experiments. The difference threshold was measured to obtain the smallest time difference necessary to displace the auditory or cutaneous sensation from 0° azimuth. Intensity was always set to equal loudness and *S*'s task was to indicate whether localization produced by a given time difference was to the left or the right of the position produced by a time difference of zero.

Method

Subjects.—Three male students were employed as *S*s for Exp. IV. Since none of the *S*s had had previous experience in auditory- or cutaneous-localization experiments, several sessions were devoted to preliminary training.

Apparatus.—The apparatus used in Exp. II and III was also used in the present experiment.

Procedure.—Difference thresholds were measured for the ears, bilateral index fingertips, ipsilateral fingertips, single index fingertip, and the volar forearm. On the forearm the vibrators were placed 10–12 cm. apart and on the single index fingertips the vibrators were 1 cm. apart. Thresholds were obtained for each *S* under each of the five conditions of stimulation. Auditory thresholds were obtained first for the three *S*s and the order of cutaneous conditions was determined for each *S* by a Latin-square design. The index-finger thresholds, however, were not included in the Latin square but were always measured last, as this portion of the experiment was an afterthought.

Stimuli with varied time differences, favoring either the right or left channel, were delivered in a random order through earphones or vibrators, and *S* was required to indicate "left" or "right" when comparing the sensation to that produced by a standard stimulus with a zero time difference. The standard stimulus was presented before the presentation of each comparison stimulus. The number of time-difference values used (ranging from large delays to the left to large delays to the right) varied between 9 and 15. From each *S* were obtained 10 judgments for each time-difference value.

Results and Discussion

Difference thresholds were determined by fitting normal curves to the functions relating probability of "right" responses to time differences. The distance along the time-difference scale between the 75% value and the PSE was measured to obtain the "right" threshold, and the "left" threshold was likewise obtained by measuring the distance between the 25% value and the PSE. The curve fitting was performed by transforming *p* values to *z* values, which were plotted as a function of time differences, and a straight line was drawn through the points by inspection.

As in Exp. II and III, the ears (mean threshold .046 msec.) were more sensitive to time differences than were the bilateral fingertips (mean threshold 4.550 msec.) or ipsilateral fingertips (mean threshold 1.095 msec.). Also in agreement with the results of Exp. II and III is the finding that the ipsilateral-cutaneous thresholds were considerably smaller than the bilateral thresholds. When the vibrators were placed 10–12 cm. apart on the forearm the thresholds (mean threshold .982 msec.) were slightly smaller than when they were on the ipsilateral fingertips, but when they were placed 1 cm. apart on the same index finger, cutaneous thresholds were found to be relatively small (mean threshold .229 msec.). The sensitivity of the ears and the single fingertip to time differences seemed to at least approach each other.

These findings are consistent with those of Exp. I, II, and III, indicating that time delays produce different effects for the ears and skin and that the magnitude of the cutaneous effects is highly dependent upon the spatial relationship between stimulated areas. It appears that the time-difference effect is a function of neural inhibition and that the amount of inhibition produced by a given time difference is related to the degree of neural interaction between stimulated areas. Békésy (1957), however, has reported time-difference effects to be the same for stimulation of the ears, thumb, finger, arm, leg, and chest, and even the bilateral fingertips (Békésy, 1958). The published descriptions of Békésy's experiments afford no obvious explanation of the difference between his findings and those reported here.

REFERENCES

Békésy, G. v. Human skin perception of traveling waves similar to those on the cochlea. *J. Acoust. Soc. Amer.*, 1955, **27**, 830–841.

Békésy, G. v. Sensations on the skin similar to directional hearing, beats, and harmonics of the ear. *J. Acoust. Soc. Amer.*, 1957, **29**, 489–501.

Békésy, G. v. Funneling in the nervous system. *J. Acoust. Soc. Amer.*, 1958, **30**, 399–412.

Rayleigh, Lord. Acoustical observations. *Phil. Mag.*, 1877, **3**, Ser. 5, 456–464.

Sivian, L. J., & White, S. D. On minimum audible sound fields. *J. Acoust. Soc. Amer.*, 1933, **4**, 288–321.

Stevens, S. S., & Newman, E. B. The localization of actual sources of sound. *Amer. J. Psychol.*, 1936, **48**, 297–306.

Woodworth, R. S., & Schlosberg, H. *Experimental psychology.* (Rev. ed.). New York: Holt, 1955.

(Received July 22, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 6, 626-631

EFFECTS OF INTERVAL BETWEEN SUCCESSIVE NUMBERS AND PATTERN IN VERBAL LEARNING

CHARLES P. THOMPSON[1]

University of Wyoming

The effect of interval between successive numbers (IBSN) and the interactions among IBSN, doublet, and doublet location in a serial list were investigated. In one experiment, 48 Ss learned a verbal maze under 1 of 2 IBSN conditions. In another experiment, 160 Ss were divided into a large and small IBSN condition for each of 4 doublet-location conditions. The lists were learned by the serial-anticipation method. The experiments demonstrated that rate of acquisition of a serial list increases as IBSN decreases. Further, the doublet, in general, acts as an anchoring point in serial-anticipation learning and the effectiveness of this anchoring doublet increases as the difficulty of the series in which it is located increases.

Using a free-responding number task, Jenkins and Sheffield (1946) found that, with the exception of a number succeeding itself, the probability of a particular number response increases with its ordinal proximity to the previous number response. These data strongly suggest that a list composed with an interval between successive numbers (IBSN) that is small (e.g., 6-5-7) and in which no number succeeds itself would be learned more rapidly than a comparable list composed with large IBSN (e.g., 2-9-4). This hypothesis is supported by the results of other research establishing the nonrandom response tendencies of *S*s (e.g., Smith, 1949; Underwood & Schulz, 1960).

In addition, serial-position pattern effects (e.g., Ernst, Hoffeld, Seidenstein, & Brogden, 1960; Ernst, Thompson, & Brodgen, 1962; Namikas, Thompson, & Brogden, 1960) should vary as a function of the IBSN characteristics of the list in which the pattern appears. For example, the effect of the doublet (two successive identical responses) in a serial-anticipation task consists of a reduction in errors at the doublet locations and is maximal at those positions which are most difficult for *S* to learn (Ernst et al., 1962). If, owing to their uniqueness, the doublet positions act as anchor points in a series, it is reasonable to assume greater effectiveness for these anchor points as the series becomes more difficult. In number lists, difficulty should vary with variations in IBSN characteristics. Therefore, the doublet effect should vary as a function of the IBSN characteristics of the list in which the doublet appears.

Two experiments were conducted to test the hypotheses stated in the previous paragraphs. The first experiment investigated rate of acquisition of a verbal maze as a function of IBSN. The second extended the IBSN results to the serial-anticipation procedure and, in addition, investigated the interactions among IBSN, doublet, and doublet location.

EXPERIMENT I

Method

Subjects.—Fifty-two undergraduate students enrolled in introductory psychology at the University of Wyoming served as *S*s. Three were discarded for exceeding the 1-hr. time limit and one was discarded because of procedural error. All *S*s were naive with respect to the experimental task.

---

[1] Now at Kansas State University.

Experimental design.—A simple randomized design with 48 *S*s divided into large and small IBSN groups was used. The *S*s were assigned at random to one of these two groups. Each *S* was required to learn a 16-unit, eight-choice verbal maze using the correction procedure. The eight choices were the numbers 10, 20, 30, 40, 50, 60, 70, and 80. The mazes were constructed so that each choice occurred twice and so that there were no repetitive patterns in the mazes. Four different mazes were used in each group as a control for other possible sequence effects. Six *S*s within each group were randomly assigned to each submaze. The range of IBSN in the large IBSN set was between 30 and 60 (e.g., 10-70-30-60-20) with the median IBSN being 40. The range of IBSN in the small IBSN set was between 10 and 40 (e.g., 70-50-10-30-40) with the median IBSN being 20.

Procedure.—Each *S* was given slightly modified instructions from Thompson, Voss, and Brogden (1957). The only modifications of these instructions were the substitution of the word "correct" for the word "forward," and the addition of the four extra numbers used in this experiment. After questions were answered, the learning session was begun. Acquisition trials were continued until *S* reached the criterion of one errorless trial. The *S*s not reaching criterion within 1 hr. were discarded. The intertrial interval was held at 30 sec. Time for each trial and each response at each serial position for each trial were recorded for each *S*.

Results

Separate analyses of variance were performed on the total-time, total-trials, and total-errors data. These analyses found the IBSN variable to be significant for the total-time measure, $F\ (1, 46) = 10.47$, $p < .01$. The means for this measure were 920 sec. and 1451 sec. for the small and large IBSN conditions, respectively. No other source of variation reached significance in these analyses. The differences in the total-trials and total-errors measures were in the expected direction but were not significant.

Discussion

The hypothesis that IBSN would be an effective variable was confirmed only for the total-time measure. Since the serial-anticipation procedure typically employs automatic pacing in which time and trials measures are identical, it was not clear whether IBSN would be an effective variable when using a paced serial-anticipation task. The second experiment was designed to answer that question and, at the same time, to investigate interactions among IBSN, doublet, and doublet location.

EXPERIMENT II[2]

Method

Subjects.—One hundred and seventy-one students enrolled in introductory psychology at the University of Wyoming served as *S*s. Eleven *S*s were discarded during the course of the experiment. Of the 11 discards, 4 were due to equipment failure, 3 to procedural error, and 4 did not complete the task. All *S*s were naive with respect to the experimental task.

Experimental design.—An orthogonal design with 160 *S*s divided into a large and small IBSN condition for each of four doublet-location conditions (locations in each of the first, middle, and last thirds of the list plus a control list with no doublet) was used. The *S*s were assigned randomly within replication to one of the eight groups. Each *S* was required to learn a 16-unit, eight-choice list using a paced serial-anticipation procedure. Four large and four small IBSN control (no doublet) lists were constructed in the same manner and using the same alternatives as the lists in Exp. I. The experimental lists were constructed, under the same restrictions, by inserting a doublet into the control lists with as few other alterations as possible. The doublet was inserted at Positions 3 and 4, 8 and 9, and 13 and 14 for the first, second, and third doublet locations, respectively. Five *S*s within each of the eight groups were randomly assigned to each of the four sublists. The range of IBSN in the large IBSN set was between 20 and 70 with the median IBSN being 40. The range of IBSN in the small IBSN set was between 10 and 40 with the median being 20.

---

[2] The author wishes to express his appreciation to Burton Routman for his assistance in the collection and analysis of the data for this experiment.

Procedure.—The instructions of Ernst et al. (1962) were modified slightly to make them appropriate to this experiment and were given by *E* to each *S*. After questions were answered, the appropriate list was presented to *S* with a Dunning animatic stripfilm projector. Each stimulus was presented for 2 sec. and all were presented successively until the end of the list was reached. The intertrial interval was 30 sec. The viewing distance was approximately 3 ft. and the stimuli were $1\frac{1}{4}$ in. high on a white background. Acquisition trials were continued until *S* reached the criterion of one errorless trial. The *S*s not reaching criterion within 1 hr. were discarded. Each response for each position in the list was recorded for each trial.

Results

Analyses of variance were performed on the total errors by serial-position and total-trials data.

The IBSN effect was significant for total trials anova, $F\ (1, 152) = 7.40$, $p < .01$. The means for the small and large IBSN conditions were 14.7 and 18.0 trials, respectively. No other source of variation reached significance in this anova.

IBSN, serial position, Serial Position × IBSN, Serial Position × Doublet Location, and Serial Position × Doublet Location × IBSN were all significant beyond the .001 level in the Total Errors × Serial-Position analysis. No other source of variation reached significance in this anova.

The significant IBSN effect shows that total errors to criterion are a function of the IBSN characteristics of the list. The means for the IBSN main effect were 84.7 and 118.6 total errors for the small and large IBSN conditions, respectively.

The significance of the serial-position term demonstrates the expected bowed serial-position error curve.

Inspection of the serial-position error curves for the large and small IBSN conditions indicated that no meaningful interpretation could be given to the significant Serial Position × IBSN interaction.

In order to evaluate adequately the interaction effects involving the doublet, the total-errors data were subjected to an experimental minus control (E − C) transformation of the type employed by Ernst et al. (1960). Separate E − C transformations were performed for the small and large IBSN conditions on the total errors at each doublet position and the two adjacent positions. These data were then analyzed to evaluate the effect of the doublet relative to the corresponding control positions as a function of doublet location and IBSN characteristics of the list.

Doublet location, serial position of the doublet (SPD), SPD × Doublet location, and SPD × Doublet Location × IBSN were all significant be-

TABLE 1

Summarized Results of the Duncan Range Test Performed on the Means of the Transformed (E-C) Total Errors for the Doublet and Adjacent Positions

Condition	L3	S2	L2	S3	S1	L1
Mean	−5.10	−3.85	−3.31	−2.79	−1.53	−.40

Note.—Means underlined by the same line do not differ significantly from one another. Means not underlined by the same line do differ significantly from one another. The letters L and S stand for the large and small IBSN conditions, respectively. The numbers 1, 2, and 3 stand for the first, second, and third doublet locations, respectively. The means are for each of the three doublet locations in the large and small IBSN conditions.

yond the .001 level in the E − C anova. In addition, the SPD × IBSN interaction was significant at beyond the .01 level and the Doublet Location × IBSN interaction was significant at beyond the .05 level. No other source of variation reached significance in this analysis.

The significance of the doublet-location term demonstrates that the difference between errors in control and experimental groups made at the doublet and adjacent positions varies as a function of the location of the doublet. The mean E − C total errors were −.97, −3.58, and −3.94 for the first, second, and third doublet locations, respectively. A Duncan range test (Duncan, 1955) performed on these means showed that the mean E − C total errors at the first doublet location differed significantly from the corresponding errors at the second and third doublet locations. The mean E − C errors at the latter locations did not differ significantly from one another.

The significant Doublet Location × IBSN interaction demonstrates that the effect of doublet location varies as a function of IBSN. A Duncan range test was performed on the appropriate E − C total-error means. The results of this test are summarized in Table 1. These results demonstrate that the doublet effect becomes greater from the first to the third doublet location in the large IBSN condition. On the other hand, the maximal effect for the small IBSN condition is found at the second doublet location with the second and third doublet locations not differing significantly from one another.

The significance of the SPD term, the IBSN × SPD interaction, and the SPD × Doublet Location interaction demonstrate the overall effect of the doublet and that this effect varies with IBSN and location within the list. These effects are presented graphically on the left-hand side of Fig. 1. The doublet effect tends to be greater for the large IBSN condition and maximal effects occur at the second and third doublet locations.

Finally, the significant triple interaction of SPD × Doublet Location × IBSN demonstrated that serial-position fluctuations as a function of doublet location are different for the large and small IBSN conditions. The serial-position data by IBSN condition for the first, second, and third doublet locations are presented on the right-hand side of Fig. 1. For the first doublet location, a greater reduction in errors at the doublet positions is found for the small IBSN condition than is found for the large IBSN condition. The reverse is true for the second and third doublet locations.

Discussion

The hypotheses which the experiment was designed to test were confirmed in part. A list with a small IBSN is learned more rapidly in terms of both total trials and total errors than is a list with large IBSN. From the work of other *E*s (e.g., Jenkins & Sheffield, 1946), it follows that this may be attributed to previously learned number habits which lead to an increase in the probability of a correct response as the ordinal proximity of that response to the previously correct response increases.

In accord with previous investigations (e.g., Ernst et al., 1962), the effect of the doublet in serial-anticipation learning was found to be a reduction in errors at the doublet and adjacent positions. The maximum reduction in errors is found at the second position within the doublet. This effect undoubtedly can be attributed to the uniqueness of the doublet in the series of numbers.

The change in the doublet effect as a function of location within the list and

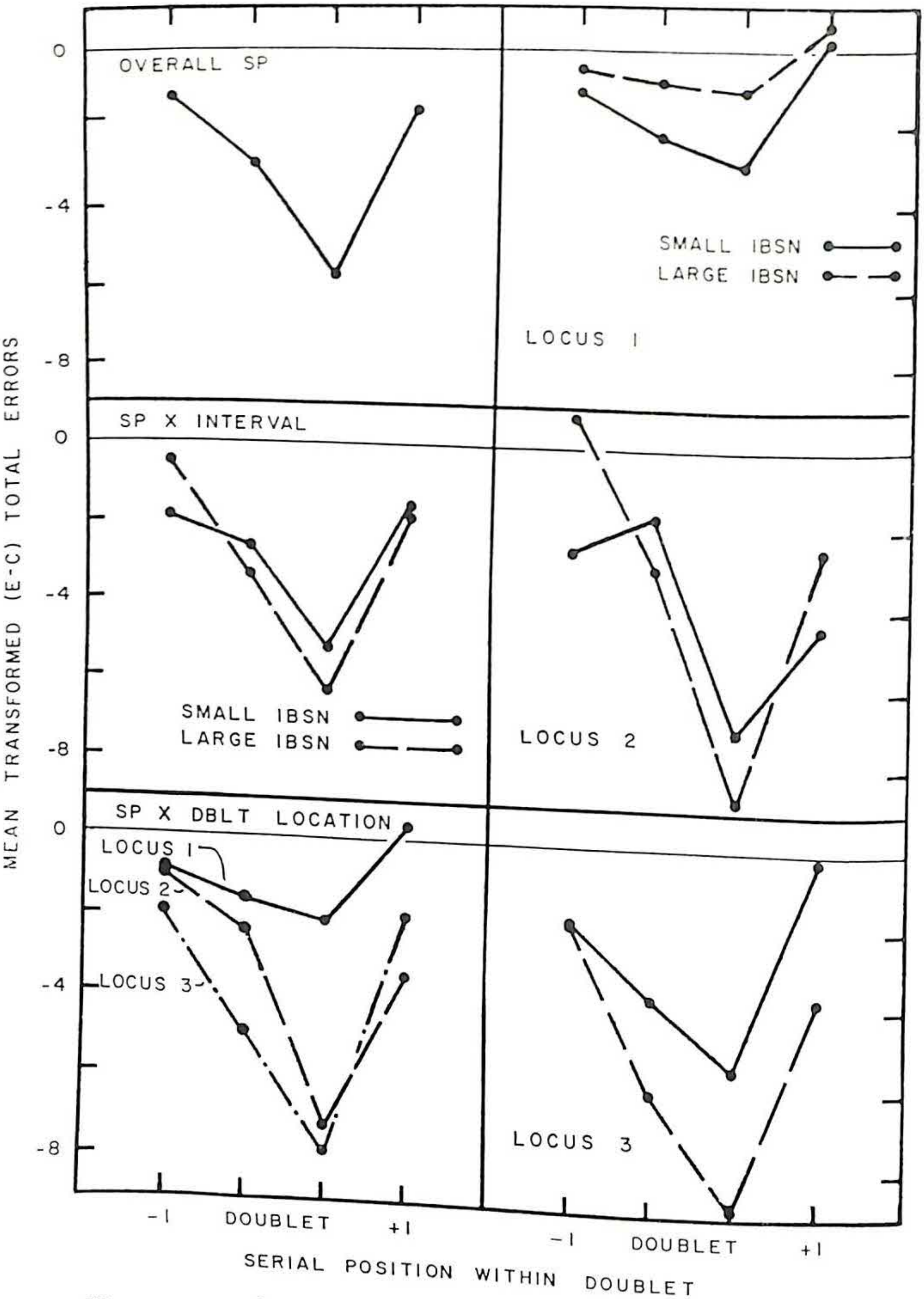


FIG. 1. The curves on the left represent, from top to bottom, the overall doublet effect, doublet effect as a function of IBSN condition, and the doublet effect as a function of doublet location. The curves on the right show the effect of the doublet as a function of IBSN condition for the first, second, and third doublet locations in the top, middle, and bottom figures, respectively. (The zero base line represents the mean total errors at the comparable control-list positions.)

IBSN condition may be ascribed to the increasing effectiveness of the pattern. Any unique positions within a serial list (e.g., the first and last positions) serve as anchoring points and produce a reduction in the number of errors made at those and adjacent positions. It is reasonable to assume that the effectiveness of any anchoring point increases as the difficulty of the items surrounding it increases. For example, a centrally located anchoring point would be of negligible value in a series which could be learned in 1 or 2 trials, but such an anchoring point would be of great value in a series requiring 20 or 30 trials for acquisition. The results of this study strongly support this hypothesis. The effect of the doublet increases with the difficulty of the list in which it is contained and with the difficulty of the portion of the list in which it is located. However, an exception exists for the first doublet location, where a greater reduction in errors is found for the smaller IBSN condition. This reversal is probably a result of the different doublet effects found for the discovery and acquisition phases of learning (Ernst et al., 1962). An increase in errors relative to the control occurs at the second position within the doublet during the discovery phase. This effect can be attributed to previously learned number habits (Thompson et al., 1957). During the acquisition phase, however, the doublet effect is reversed. This reversal can be attributed to the uniqueness of the doublet in the serial list (Thompson, 1962). Hence, for any procedure in which the duration of the discovery phase is relatively long (e.g., the verbal maze), the overall effect of the doublet is an increase in errors at the second position within the doublet. On the other hand, this effect is reversed for any procedure with a relatively short discovery-phase duration (e.g., serial anticipation). The contradictory results in this study occurred at Positions 3 and 4, which are rapidly learned. For these positions, the duration of the discovery phase, relative to the duration of the acquisition phase, is probably longer for the larger IBSN condition. A reversal of the doublet effect as a function of IBSN condition might consequently be expected. However, it is clear that additional research is necessary to establish the relationships among IBSN characteristics, discovery-phase duration, and the effect of the doublet.

To summarize, the results of this study demonstrate that, in general, the doublet acts as an anchoring point in a list learned by the serial-anticipation method and the effectiveness of this anchoring doublet increases with the difficulty of the series in which it is located.

REFERENCES

DUNCAN, D. B. Multiple range and multiple *F* tests. *Biometrics*, 1955, **11**, 1–42.

ERNST, R. L., HOFFELD, D. R., SEIDENSTEIN, S., & BROGDEN, W. J. Relation of serial position errors to doublet and split-doublet location in verbal maze pattern. *J. exp. Psychol.*, 1960, **59**, 94–103.

ERNST, R. L., THOMPSON, C. P., & BROGDEN, W. J. Effect of pattern and pleonasm location in serial lists upon acquisition and serial position errors. *J. exp. Psychol.*, 1962, **64**, 151–157.

JENKINS, W. O., & SHEFFIELD, F. D. Rehearsal and guessing habits as sources of the "spread of effect." *J. exp. Psychol.*, 1946, **36**, 316–330.

NAMIKAS, G., THOMPSON, C. P., & BROGDEN, W. J. Effect of triplet and quadruplicate location in verbal maze patterns upon serial position errors. *J. exp. Psychol.*, 1960, **59**, 383–390.

SMITH, M. H. Spread of effect is the spurious result of non-random response tendencies. *J. exp. Psychol.*, 1949, **39**, 355–368.

THOMPSON, C. P. Variation in the doublet effect during acquisition and relearning as a function of number of doublets in the maze and length of the retention interval. Unpublished doctoral dissertation, University of Wisconsin, 1962.

THOMPSON, R. F., VOSS, J. F., & BROGDEN, W. J. Effect of pattern variation upon verbal maze learning. *J. exp. Psychol.*, 1957, **54**, 253–258.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning.* Philadelphia: Lippincott, 1960.

(Received July 23, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 6, 632–635

PREFERENCE FOR A POSITIVE EVALUATIVE RESPONSE IN CONCEPT LEARNING

RAMON J. RHINE

University of California, Riverside

A comparison was made between 26 Ss learning a positive evaluative concept and 31 Ss learning a negative concept. Concepts were developed by selectively reinforcing responses to 50 trait words judged in preliminary tests as desirable or undesirable. Resistance to concept-attitude change was determined by reversing the pattern of reinforcement for 54 further traits. Scores were also obtained on authoritarian attitudes predictive of cognitive rigidity. The positive concept was more readily learned and more resistant to change. Greater resistance to change was also found for Ss high in authoritarianism than for low scorers. The results were interpreted as indicating that studies of evaluative concepts are applicable to the understanding of natural attitudes, despite difficulties in establishing a neutral base line for developing evaluative concepts.

A concept attitude has been defined as a concept with an evaluative dimension, such as *good* or *bad* (Rhine, 1958). This approach, which has been used by DiVesta (1961), Rhine (1960), Rhine and Silun (1958), and Walker (1962), makes available the laboratory techniques of concept formation for studying attitudes.

Two implicit assumptions are made in using concept-formation techniques for studying concept attitudes. First, it is assumed either that *S*s start from a position of neutrality toward the group about which a concept attitude is being developed or that related dispositions can be identified and controlled. Second, laboratory findings on concept attitudes are assumed applicable to natural attitudes. If *S*s start from a neutral point, comparison of a group learning positive with one learning negative concept attitudes should produce differences neither on concept-attitude development nor on resistance to change. If a concept attitude is a reasonable analog of natural attitudes, then it would be expected that the cognitive rigidity characteristic of *S*s with authoritarian attitudes (e.g., Rokeach, 1948; Rokeach, 1960) would result in resistance to concept-attitude change.

METHOD

Subjects and materials.—Fifty-seven college sophomores participated in the experiment. Concept-attitude development, strength, and change were determined from responses to descriptive traits printed on 4 × 6 in. cards. The traits were selected on the basis of preliminary tests with a different group of 100 *S*s who rated the traits with 90% or greater frequency as either desirable or undesirable. Items from the Fascist, Anti-Semitic, and Ethnocentric attitude scales of the *Authoritarian Personality* (Adorno, Frenkel-Brunswick, Levinson, & Sanford, 1950) were administered as a measure of a natural-attitude cluster.

Procedure.—As in previous studies of concept attitudes (e.g., Rhine and Silun, 1958), concept attitudes were developed by instructing *S*s to predict whether traits shown one at a time are characteristic or not characteristic of an unnamed fictional group of people. A positive concept-attitude was developed in a group of 26 *S*s by *E* saying after each of *S*'s responses to desirable traits that the trait is "characteristic" of the people and by saying "not characteristic" after those traits rated undesirable. For a second group, containing 31 *S*s, a negative concept attitude was developed by *E* responding "characteristic" to negatively rated traits and "not characteristic" to positive traits.

For concept-attitude development, 25 positive and 25 negative traits were presented together in random sequence. Immediately following presentation of these 50 traits, concept-attitude strength was measured by obtaining responses to 25 different, randomly sequenced traits with *E* no longer responding. Concept-attitude change was then induced by presenting 54 further traits in random order, but with *E* responding in a manner opposite from that of the development session.

All traits were presented in one session. Items from the Fascist, Anti-Semitic, and Ethnocentric scales, consisting respectively of 14, 16, and 20 items selected for their relatively high discriminability, were administered together in a separate session. These 50 items are referred to as the FASE scale. The three subscales are positively intercorrelated (Adorno et al., 1950). The total FASE scale, which provides better reliability than a smaller subscale, was used as an index of a natural attitudinal complex. The FASE scale was taken by 23 *S*s from the positive-development group and 29 *S*s from the negative group.

RESULTS

For both development and change, an *S*'s score is the sum of his correct predictions of *E*'s responses. The strength score is the sum of responses that are consistent with the concept-attitude dimension, positive or negative, that was stressed in the development session.

The primary measure upon which the results are based is *S*'s change score minus his development score. If *S*s are disposed to respond positively, scores on positive concept-attitude development should be higher than scores on negative development. The *S*s should also tend to resist change from a positive to a negative concept attitude and should switch readily from negative to positive. Consequently, the difference between *S*s' development and change scores will be more sensitive to a positive disposition than either of these scores alone because this difference takes into account both effects at once. Similarly, a negative disposition should also be detected by the difference between development and change scores. A trend in neither direction should produce difference scores which are the same for groups learning either a positive or a negative concept attitude.

TABLE 1

MEANS, *SD*s, AND *t*'S FOR MEASURES OF POSITIVE AND NEGATIVE CONCEPT ATTITUDES

Concept-Attitude Measure		Group		*t*
		Positive Development	Negative Development	
Development	*M*	40.58	36.54	1.74
	SD	3.49	7.06	
Strength	*M*	22.77	22.06	.76
	SD	3.35	3.59	
Change	*M*	36.12	40.35	1.57
	SD	11.53	8.81	
Change Minus Development	*M*	−4.46	3.81	8.35***
	SD	3.97	3.56	

***$p < .001$.

Means and *SD*s for the concept-attitude measures are presented in Table 1 for the positive- and negative-development groups. Also shown are the results of *t* tests comparing the positive and negative groups on each concept-attitude score. The means suggest that positive development is slightly greater than negative and is a little more difficult to overcome. The combined effects of the development and change sessions is reflected in the difference scores, for which the means are significantly different beyond the .001 level. The strength means, which are almost identical, help emphasize this result, since they show that both groups reach the same level before the change session begins.

The FASE scores were split into two sets consisting of 16 cases from the high end of the distribution and 16 from the low end. These scores were used to test for differences on the concept-attitude measures associated with a natural-attitude cluster. The 16 high scorers were made up of the 8 highest FASE scorers from the positive-development group and the 8 highest from the negative group, and the 16 lows were similarly selected.

Each of the three subscales of the FASE scale were scored so they could vary from one (disagree) to seven (agree) with four as the midpoint (neither agree nor disagree). The total FASE score was found by adding the individual score of each *S* on the three subscales; thus, the FASE score could vary between 3 and 21. The eight high scores of the positive-development group ranged from 11.09 to 15.41; the highs from the negative group ranged from 10.57 to 14.48; the positive lows from 4.48 to 8.82; and the negative lows from 5.13 to 7.78.

The means and *SD*s of the concept-attitude measures for the high and low FASE groups are shown in Table 2. Also shown are *t*'s for the difference between the FASE groups.

TABLE 2

MEANS, *SD*s, AND *t*'s COMPARING HIGH VS. LOW FASE SCORERS ON CONCEPT-ATTITUDE MEASURES

Concept-Attitude Measure		FASE Score		*t*
		Low	High	
Development	*M*	37.75	38.31	.22
	SD	8.69	5.49	
Strength	*M*	22.75	22.31	.35
	SD	3.73	3.62	
Change	*M*	43.94	32.06	3.17**
	SD	6.88	13.04	
Change Minus Development	*M*	6.19	−6.25	2.86**
	SD	8.74	5.50	

** $p < .01$.

In both the change and difference scores, the high and low FASE groups differ beyond the .01 level: The more authoritarian the person's orientation, the more he resists changing his concept attitude. The method of selecting high and low FASE scores, by neutralizing the tendency toward a positive response, makes the development means in Table 2 closer to each other than those in Table 1, where the difference of 4.04 falls statistically between the .10 and .05 levels. The strength scores of Table 2, as in Table 1, are identical. FASE is significantly related neither to the learning measure (concept-attitude development) nor the leveling measure (concept-attitude strength). This indicates that the difference in resistance to change is not due to a difference between high and low FASE scorers in the ability to form concepts.

Individual analysis of the F, A-S, and E subscales produce results approximating those shown in Table 2. All differences between highs and lows are in the same direction as those in Table 2, and none of the differences for strength and development reaches the .05 level. For the change and difference scores, the *t*'s for the E scale are significant beyond the .01 level, *t*'s for the F scale are significant beyond .05, but the *t*'s for the A-S scale do not reach the .05 level.

DISCUSSION

The data of Table 1 indicate that college sophomores from a state university are biased to respond positively toward an unknown group of people. The strength scores suggest that the negative and positive dimensions are both learned to the same degree before concept-attitude change begins. Nevertheless, the trend in the change session is for *S*'s with the positive concept attitude to resist change more than those having originally developed a negative concept attitude. The quite significant difference between positive- and negative-development groups on change minus development indicates a nonspecific, evaluative disposition which determines in part the course of concept-attitude development and change. Identification of this effect makes possible its control when testing hypotheses about attitudes using evaluative concepts.

Experience with natural attitudes and studies of rigidity and dogmatism lead to the expectation that authoritarian *S*s will be more resistant to concept-attitude change than less authoritarian *S*s. This expectation is confirmed by the significant difference between the scores of high- and low-authoritarian *S*s in resistance to concept-attitude change. Since a prediction based on studies of natural attitudes is confirmed, the results of the FASE test lend support to the view that laboratory concept attitudes are a reasonable analog of real attitudes. Also supported is the commonly made assumption that the relationship between measures of authoritarianism and rigidity in perceptual and problem-solving tasks applies also in the realm of attitudes.

REFERENCES

Adorno, T. W., Frenkel-Brunswik, E., Levinson, D. J., & Sanford, R. N. *The authoritarian personality.* New York: Harper, 1950.

DiVesta, F. J. Contrast effects in the verbal conditioning of meaning. *J. exp. Psychol.*, 1961, **62**, 535–544.

Rhine, R. J. A concept-formation approach to attitude acquisition. *Psychol. Rev.*, 1958, **65**, 362–370.

Rhine, R. J. The effect of peer group influence upon concept-attitude development and change. *J. soc. Psychol.*, 1960, **51**, 173–179.

Rhine, R. J., & Silun, B. A. Acquisition and change of a concept-attitude as a function of consistency of reinforcement. *J. exp. Psychol.*, 1958, **55**, 524–528.

Rokeach, M. Generalized mental rigidity as a factor in ethnocentrism. *J. abnorm. soc. Psychol.*, 1948, **43**, 259–278.

Rokeach, M. *The open and closed mind.* New York: Basic Books, 1960.

Walker, L. A concept-formation analogue of attitude development. Unpublished doctoral dissertation, Pennsylvania State University, 1959.

(Received August 10, 1964)

Journal of Experimental Psychology
1965, Vol. 70, No. 6, 636

Manuscripts Accepted for Publication in the *Journal of Experimental Psychology*

The titles and authors of accepted papers are listed here following Supplementary Reports. This program is being supported on an experimental basis by the APA Project on Scientific Information Exchange in Psychology, and the outcome of this trial is presently being evaluated.

This listing plus those published in the preceding 1965 issues represents the entire backlog of manuscripts accepted by this journal. Such listing is intended to allow readers to become aware of research many months in advance of journal publication.

Atmosphere and Conversion Errors in Syllogistic Reasoning: Miles E. Simpson and Donald M. Johnson*: Department of Psychology, Michigan State University, East Lansing, Michigan.

Adults and the Intermediate Size Problem: Michael D. Zeiler* and Jo Lang: Psychological Laboratory, Pendleton Hall, Wellesley College, Wellesley, Massachusetts 02181.

Paired-Associate Learning and the Timing of Arousal: D. E. Berlyne,* Donna M. Borsa, Jane H. Hamacher, and Isolde D. V. Koenig: Department of Psychology, University of Toronto, Toronto 5, Canada.

Stimulus Generalization of a Positive Conditioned Reinforcer IV: Concurrent Generalization of Reinforcing and Discriminative Stimulus Functions Following Fixed Interval Training: David R. Thomas* and Donald V. DeRosa: Department of Psychology, Kent State University, Kent, Ohio 44240.

Symmetry in Paired Associates: Harald R. Leuba*: Department of Psychology, Johns Hopkins University, Baltimore, Maryland 21218.

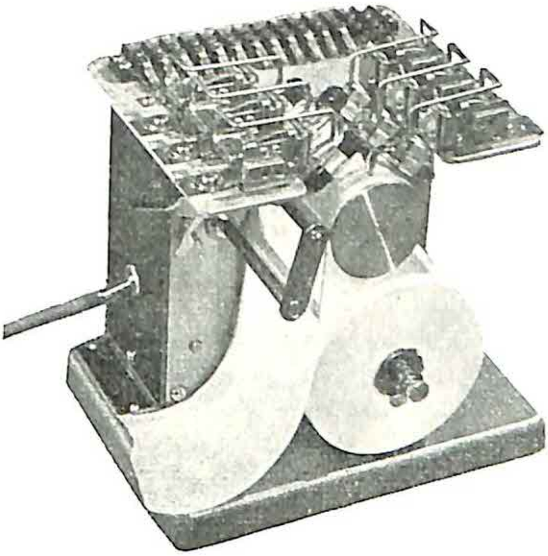
Stimulus Specificity: Nonreward: E. J. Capaldi*: Department of Psychology, Mezes Hall 211, University of Texas, Austin 12, Texas.

Remembrance of Lines Past: Anita R. Cunitz* and Bruce M. Ross: Institute for Behavioral Research, 2426 Linden Lane, Silver Spring, Maryland.

Magnitude of Gravitoinertial Force, an Independent Variable in Egocentric Visual Localization of the Horizontal: Earl F. Miller, II* and Ashton Graybiel: Physiological Optics Branch, United States Naval Aviation Medical Center, Pensacola, Florida 32512.

---

* Asterisk indicates author for whom address is supplied.



- one of the most

useful

instruments in any lab is our

- **LOW COST**

• Event Recorder

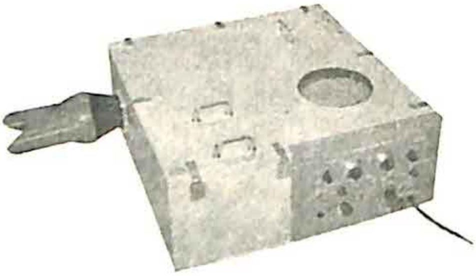
- A FULL SIZE, EASY LOADING, MULTISPEED INK WRITING INSTRUMENT OF DURABLE CONSTRUCTION
- WRITES A CLEAR LEGIBLE RECORD ON INEXPENSIVE PAPER (700 FOOT ROLL COSTS ONLY 1.00)
- SPECIFICALLY DESIGNED TO MEET THE NEEDS OF RESEARCHERS AND STUDENTS. THE RECORDER IS NOT ENCASED. ALL PARTS AND CONNECTORS ARE READILY ACCESSIBLE AND INTERCHANGEABLE
- ALL COMPONENTS ARE RATED FOR CONTINUOUS DUTY. PEN COILS WILL NOT BURN OUT
- AVAILABLE IN 4 AND 6 PEN MODELS
- FOR COMPLETE DETAILS WRITE OR PHONE

Ralph Gerbrands Company

Scientific Instruments

8 BECK RD., ARLINGTON 74, MASS. • TEL. 617-648-6415

NEW . . . FROM SCIENTIFIC PROTOTYPE



A LOW COST · HIGH PERFORMANCE FULL CAPABILITY · TWO CHANNEL TACHISTOSCOPE

A new tool for the researcher. A two-channel Tachistoscope for use where all the extraordinary features of the world renown 3-channel Scientific Prototype GB Tachistoscope's are not required. The 2-channel Model 800E, however, . . . at a far lower price . . . offers all the accuracy, proven features, and the quality that made the Model GB famous.

Much of the design and capability of the Model 800E is based on the suggestions and information provided by researchers and teachers in the field.

Thus the following is just a few of the areas that can be investigated:

BASIC VISUAL PSYCHOPHYSICS—a. Area-time relationships (Ricco's and Piper's laws) b. Intensity-time relationships (Bloch's and Pieron's laws) c. Dark and Light Adaption d. Relative and Absolute Brightness Thresholds e. Visual Acuity f. Broca-Sulzer Phenomenon. PERCEPTION—a. Perceptual Defense b. Duration Thresholds for Symbols, Words, Numbers, etc. c. Figure-Ground Phenomena d. Contour Perception e. Figural After-Effects.

FEATURES: Accepts standard 4" × 5" stimulus cards or transparencies. ● Electronic system contains two separate time interval generators. ● Range 1.0 msec to 11 sec. in four overlapping direct reading ranges. ● Accuracy guaranteed ± 2% of any setting. ● Electronic intensity controls over approximately 1 log unit. ● Field locks allow direct operator intervention without disturbing logic. ● Logic provided to allow continuous alternating of fields, overlapping (competitive) exposures and the use of one field as a pre and post exposure adaption and fixation field, while other field is used for stimulus exposure. ● Blower cooled electronics for minimum temperature rise, maximum reliability. ● Jack provided for remote firing of unit. Complete logic control for each timer, plus inhibit logic, plus manual field lock switches, and a solid state pulse generator which may be controlled by remote switch closure, provide maximum flexibility. ● Two lamps per field—rise and fall times guaranteed under 0.01 milliseconds, laminated front surface mirror, with magnesium fluoride anti-reflection back coating. Adjustable mirror mount for field alignment, four internal masks with adjustable edges for stimulus framing and elimination of internal cues.

All power supply, logic, lamp current control and timer semiconductors, are silicon for maximum reliability and negligible temperature dependence. Logic and lamp current reference power supplies are regulated. Timing and lamp current are independent of line voltage.

Electronic intensity controls plus filter holders for neutral density filters or transparent stimuli (4" × 5"). Reversible holders for 4" × 5" stimulus cards.

Special Features: High voltage interlock, Low noise forced ventilation, Fitted viewing hood, Compact self-contained construction. No inter-connecting cables required.

Priced $845, f.o.b., New York City. Delivery from stock.

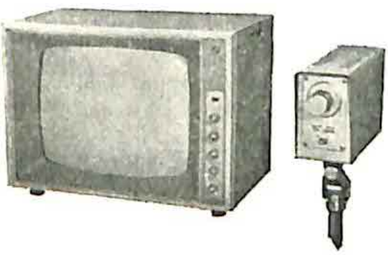
**MANUAL AVAILABLE: "Laboratory Manual in Vision for 2-channel Tachistoscopes"
$3.00 plus postage.**

SCIENTIFIC PROTOTYPE MFG., CORP.

Behavior Analysis Apparatus . . . reed relay control equipment cages, pellet dispensers, soundproof chambers, recorders, noise generators, shock sources, timers, switches, digital test equipment and special systems.

615 W. 131st Street
New York, New York 10027
Telephone: (212) 368–6000

SCIENTIFIC PROTOTYPE



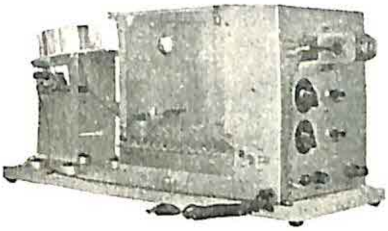
LOW COST, CLOSED CIRCUIT TV

The quality and extremely low cost of this new closed circuit TV camera opens many new avenues of research to the clinician and researcher.

Two outputs are standard. One is the usual 4 mc output to a closed circuit receiver. The other is a standard Channel 6 output that can be fed to any commercial TV receiver.

The camera is supplied with a standard F/1.4, 25 mm C mount lens that will take pictures under less than normal room illumination.

A full selection of accessories is available so that the camera can be made to perform many versatile programs. They are wide angle, manual and automatic zoom and telephoto lenses, tripod adapters, cables, line amplifiers, and transformers. Delivery is off-the-shelf. Price only $450.00. For HC-13 TV camera, 16″ VM-163 Monitor $300.00. Prices F.O.B. New York.



STUDENT LABORATORY MODEL A-105 RODENT TEST CAGE

This unit was designed by psychologists to provide a self-contained unit which is compact in size, flexible in use, versatile in concept. The rodent box is approximately 9″ × 8⅝″ × 9″.

A 2-amp DC power supply is mounted piggyback on the cage as shown. The power supply has the following controls: 1. 24 VDC power supply output. 2. A manually resettable three digit Veeder Root counter, (supplied wired to count lever presses). 3. A pulsed DC output from the pedal (supplied wired to the pellet feeder). 4. Two independent lever switches, one for shock, the other for cue light. These switches are spring loaded in one direction for manual momentary operation. The other position leaves shock and/or cue light on. 5. A light intensity control (3 position switch). 6. A potentiometer for shock intensity, control. 7. A safety shock switch (next to the prewired grids). 8. Power switch with pilot light and fuse. An automatic pellet feeder on shock mounts (at the opposite end of the cage). The cage, power supply, and feeder are mounted on ¾″ plywood base with rubber feet. Price: $259.00, f.o.b., New York City. A liquid feeder can be added to the unit for an additional $90.00. However, a liquid feeder may be substituted for the pellet feeder for an additional $10.00. (Use Model #108 when ordering.) Modifications to panels or wiring can be made at a slight increase in cost.

SCIENTIFIC PROTOTYPE MFG., CORP.

Behavior Analysis Apparatus . . . tachistoscopes, reed relay equipment, rodent test cages, pellet dispensers, soundproof chambers, recorders, noise generators, shock sources, timers, switches, digital test equipment and special systems.

615 W. 131st Street **New York, New York 10027**

Telephone: (212) 368–6000

Academic Press

CURRENT BOOKS IN YOUR FIELD

Categories of Human Learning

edited by *A. W. Melton*

Presents a study of the most promising topics for the advancement of our understanding of human learning and the interrelationships of types of learning.

1964, 356 pp., $8.50

Paired-Associates Learning

THE ROLE OF MEANINGFULNESS, SIMILARITY, AND FAMILIARIZATION

by *Albert E. Goss* and *Calvin F. Nodine*

CONTENTS: Introduction. Scalings of Meaningfulness. Meaningfulness in Paired-Associates Learning. Theoretical Analyses. Similarity. Familiarization. Meaning and Association. Summary, Significance and Suggestions. Appendix. *Author Index—Subject Index.*

1965, 358 pp., $13.50

Deranged Memory

A PSYCHONOMIC STUDY OF THE AMNESIC SYNDROME

by *George A. Talland*

An account of the most extensive investigation ever conducted with a group of patients in the chronic state of Korsakoff's Disease. Designed to test specific hypotheses about disturbances characteristic of the amnesic syndrome, and, indirectly, the processes involved in normal memory function.

1965, 356 pp., $9.50

For subscription information please write to Journal Subscription Department.

Journal of Verbal Learning and Verbal Behavior

edited by *Leo Postman*

Journal of Mathematical Psychology

edited by *Richard C. Atkinson*

Journal of Experimental Research in Personality

edited by *I. E. Farber*

Academic Press NEW YORK AND LONDON • 111 FIFTH AVENUE, NEW YORK, N. Y. 10003 • BERKELEY SQUARE HOUSE, LONDON W.1

MATHEMATICS · PHYSICS · CHEMISTRY · BIOLOGICAL SCIENCES · MEDICAL RESEARCH · SPACE SCIENCES · ENGINEERING · PSYCHOLOGY